

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: Thomas, Winston J.  
Drayna, Dennis T.  
Feder, John N.  
Gnirke, Andreas  
Ruddy, David  
Tsuchihashi, Zenta  
Wolff, Roger K.
- (ii) TITLE OF INVENTION: PLASMIDS COMPRISING NUCLEIC ACIDS FROM THE  
HEREDITARY HEOMCHROMATOSIS GENE
- (iii) NUMBER OF SEQUENCES: 76
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Jones Day
  - (B) STREET: 222 East 41<sup>st</sup> Street
  - (C) CITY: New York
  - (D) STATE: New York
  - (E) COUNTRY: USA
  - (F) ZIP: 10017
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: Windows 95
  - (D) SOFTWARE: FastSEQ for Windows Version 2.0b
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 09/497,957
  - (B) FILING DATE: 04-FEB-2000
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/834,497
  - (B) FILING DATE: 04-APR-1997
  - (C) CLASSIFICATION:
- (viii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/652,265
  - (B) FILING DATE: 23-MAY-1996
  - (C) CLASSIFICATION:
- (ix) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/632,673
  - (B) FILING DATE: 16-APR-1996
  - (C) CLASSIFICATION:
- (x) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/630,912
  - (B) FILING DATE: 04-APR-1996
  - (C) CLASSIFICATION:
- (xi) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: George, Nikolaos C.
  - (B) REGISTRATION NUMBER: 39,201

(C) REFERENCE/DOCKET NUMBER: 8907-087-999

(xii) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 650-739-3939

(B) TELEFAX: 650-739-3900

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10825 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: join(361..436, 3762..4025, 4235..4510, 5606..5881,  
6040..6153, 7107..7147)

(D) OTHER INFORMATION: /product= "Hereditary Hemochromatosis  
(HH) protein"  
/note= "Normal or wild-type (unaffected)  
Hereditary Hemochromatosis (HH) gene  
allele"

(ix) FEATURE:

(A) NAME/KEY: -

(B) LOCATION: 140..7319

(D) OTHER INFORMATION: /note= "start and stop positions for  
normal or wild-type (unaffected) allele  
cDNA (SEQ ID NO:9)"

(ix) FEATURE:

(A) NAME/KEY: -

(B) LOCATION: 3852..3891

(D) OTHER INFORMATION: /note= "start and stop positions for  
normal or wild-type (unaffected) genomic  
sequence surrounding variant for 24d2(C)  
allele (SEQ ID NO:41)"

(ix) FEATURE:

(A) NAME/KEY: -

(B) LOCATION: 5507..6023

(D) OTHER INFORMATION: /note= "start and stop positions for  
normal or wild-type (unaffected) genomic  
sequence surrounding variant for 24d1(G)  
allele (SEQ ID NO:20)"

(ix) FEATURE:

(A) NAME/KEY: allele

(B) LOCATION: replace(3872, "c")

(D) OTHER INFORMATION: /phenotype= "normal or wild-type  
(unaffected)"  
/label= 24d2

(ix) FEATURE:

(A) NAME/KEY: allele

(B) LOCATION: replace(3878, "a")  
(D) OTHER INFORMATION: /phenotype= "normal or wild-type  
(unaffected)"  
/label= 24d7

## (ix) FEATURE:

(A) NAME/KEY: allele  
(B) LOCATION: replace(5834, "g")  
(D) OTHER INFORMATION: /phenotype= "normal or wild-type  
(unaffected)"  
/label= 24d1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TCTAAGGTTG AGATAAAATT TTAAATGTA TGATTGAATT TTGAAAATCA TAAATATTTA	60
AATATCTAAA GTTCAGATCA GAACATTGCG AAGCTACTTT CCCCAATCAA CAACACCCCT	120
TCAGGATTTA AAAACCAAGG GGGACACTGG ATCACCTAGT GTTTCACAAG CAGGTACCTT	180
CTGCTGTAGG AGAGAGAGAA CTAAAGTTCT GAAAGACCTG TTGCTTTTCA CCAGGAAGTT	240
TTACTGGGCA TCTCCTGAGC CTAGGCAATA GCTGTAGGGT GACTTCTGGA GCCATCCCCG	300
TTTCCCCGCC CCCCAAAAGA AGCGGAGATT TAACGGGGAC GTGCGGCCAG AGCTGGGGAA	360
ATG GGC CCG CGA GCC AGG CCG GCG CTT CTC CTC CTG ATG CTT TTG CAG	408
Met Gly Pro Arg Ala Arg Pro Ala Leu Leu Leu Leu Met Leu Leu Gln	
1 5 10 15	
ACC GCG GTC CTG CAG GGG CGC TTG CTG C GTGAGTCCGA GGGCTGCGGG	456
Thr Ala Val Leu Gln Gly Arg Leu Leu	
20 25	
CGAACTAGGG GCGCGGCGGG GGTGGAAAAA TCGAAACTAG CTTTTTCTTT GCGCTTGGA	516
GTTTGCTAAC TTTGGAGGAC CTGCTCAACC CTATCCGCAA GCCCCTCTCC CTACTTTCTG	576
CGTCCAGACC CCGTGAGGGA GTGCCTACCA CTGAACTGCA GATAGGGGTC CCTCGCCCCA	636
GGACCTGCCC CCTCCCCCGG CTGTCCCGGC TCTGCGGAGT GACTTTTGGA ACCGCCCCACT	696
CCCTTCCCCC AACTAGAATG CTTTTAAATA AATCTCGTAG TTCCTCACTT GAGCTGAGCT	756
AAGCCTGGGG CTCCTTGAAC CTGGAACTCG GGTATTATTC CAATGTCAGC TGTGCAGTTT	816
TTTCCCAGT CATCTCCAAA CAGGAAGTTC TTCCCTGAGT GCTTGCCGAG AAGGCTGAGC	876
AAACCCACAG CAGGATCCGC ACGGGGTTTC CACCTCAGAA CGAATGCGTT GGGCGGTGGG	936
GGCGCGAAAG AGTGGCGTTG GGGATCTGAA TTCTTCACCA TTCCACCCAC TTTTGGTGAG	996
ACCTGGGGTG GAGGTCTCTA GGGTGGGAGG CTCCTGAGAG AGGCCTACCT CGGGCCTTTC	1056
CCCACTCTTG GCAATTGTTC TTTTGCCCTGG AAAATTAAGT ATATGTTAGT TTTGAACGTT	1116
TGAAGTGAAC AATTCTCTTT TCGGCTAGGC TTTATTGATT TGCAATGTGC TGTGTAATTA	1176
AGAGGCCTCT CTACAAAGTA CTGATAATGA ACATGTAAGC AATGCACTCA CTTCTAAGTT	1236

ACATTCATAT CTGATCTTAT TTGATTTTCA CTAGGCATAG GGAGGTAGGA GCTAATAATA	1296
CGTTTATTTT ACTAGAAGTT AACTGGAATT CAGATTATAT AACTCTTTTC AGGTTACAAA	1356
GAACATAAAT AATCTGGTTT TCTGATGTTA TTTCAAGTAC TACAGCTGCT TCTAATCTTA	1416
GTTGACAGTG ATTTTGGCCCT GTAGTGTAGC ACASTGTTCT GTGGGTCACA CGCCGGCCTC	1476
AGCACAGCAC TTTGAGTTT GGTACTACGT GTATCCACAT TTTACACATG ACAAGAATGA	1536
GGCATGGCAC GGCCTGCTTC CTGGCAAATT TATTCAATGG TACACTGGGC TTTGGTGGCA	1596
GAGCTCATGT CTCCACTTCA TAGCTATGAT TCTTAAACAT CACACTGCAT TAGAGGTTGA	1656
ATAATAAAAT TTCATGTTGA GCAGAAATAT TCATTGTTTA CAAGTGTAAG TGAGTCCCAG	1716
CCATGTGTTG CACTGTTCAA GCCCCAAGGG AGAGAGCAGG GAAACAAGTC TTTACCCTTT	1776
GATATTTTGC ATTCTAGTGG GAGAGATGAC AATAAGCAA TGAGCAGAAA GATATACAAC	1836
ATCAGGAAAT CATGGGTGTT GTGAGAAGCA GAGAAGTCAG GGCAAGTCAC TCTGGGGCTG	1896
ACACTTGAGC AGAGACATGA AGGAAATAAG AATGATATTG ACTGGGAGCA GTATTTCCCA	1956
GGCAAACCTGA GTGGGCCTGG CAACTTGGAT TAAAAAGCGG GTTTTCTCAG CACTACTCAT	2016
GTGTGTGTGT GTGGGGGGGG GGGGCGGCGT GGGGGTGGGA AGGGGGACTA CCATCTGCAT	2076
GTAGGATGTC TAGCAGTATC CTGTCCTCCC TACTCACTAG GTGCTAGGAG CACTCCCCCA	2136
GTCTTGACAA CCAAAAATGT CTCTAAACTT TGCCACATGT CACCTAGTAG ACAAACCTCT	2196
GGTTAAGAAG CTCGGGTTGA AAAAAATAAA CAAGTAGTGC TGGGGAGTAG AGGCCAAGAA	2256
GTAGGTAATG GGCTCAGAAG AGGAGCCACA AACAAGGTTG TGCAGGCGCC TGTAGGCTGT	2316
GGTGTGAATT CTAGCCAAGG AGTAACAGTG ATCTGTCACA GGCTTTTAAA AGATTGCTCT	2376
GGCTGCTATG TGGAAAGCAG AATGAAGGGA GCAACAGTAA AAGCAGGGAG CCCAGCCAGG	2436
AAGCTGTTAC ACAGTCCAGG CAAGAGGTAG TGGAGTGGGC TGGGTGGGAA CAGAAAAGGG	2496
AGTGACAAAC CATTTCTCTC TGAATATATT CTGAAGGAAG TTGCTGAAGG ATTCTATGTT	2556
GTGTGAGAGA AAGAGAAGAA TTGGCTGGGT GTAGTAGCTC ATGCCAAGGA GGAGGCCAAG	2616
GAGAGCAGAT TCCTGAGCTC AGGAGTTCAG GACCAGCCTG GGCAACACAG CAAAACCCCT	2676
TCTCTACAAA AAATACAAA ATTAGCTGGG TGTGGTGGCA TGCACCTGTG ATCCTAGCTA	2736
CTCGGGAGGC TGAGGTGGAG GGTATTGCTT GAGCCCAGGA AGTTGAGGCT GCAGTGAGCC	2796
ATGACTGTGC CACTGTACTT CAGCCTAGGT GACAGAGCAA GACCCTGTCT CCCCTGACCC	2856
CCTGAAAAAG AGAAGAGTTA AAGTTGACTT TGTTCTTTAT TTTAATTTTA TTGGCCTGAG	2916
CAGTGGGGTA ATTGGCAATG CCATTTCTGA GATGGTGAAG GCAGAGGAAA GAGCAGTTTG	2976
GGGTAAATCA AGGATCTGCA TTTGGGACAT GTTAAGTTTG AGATTCCAGT CAGGCTTCCA	3036
AGTGGTGAGG CCACATAGGC AGTTCAGTGT AAGAATTCAG GACCAAGGCT GGGCACGGTG	3096

GCTCACTTCT GTAATCCCAG CACTTTGGTG GCTGAGGCAG GTAGATCATT TGAGGTCAGG 3156  
 AGTTTGAGAC AAGCTTGGCC AACATGGTGA AACCCCATGT CTAATAAAAA TACAAAAATT 3216  
 AGCCTGGTGT GGTGGCGCAC GCCTATAGTC CCAGGTTTTTC AGGAGGCTTA GGTAGGAGAA 3276  
 TCCCTTGAAC CCAGGAGGTG CAGGTTGCAG TGAGCTGAGA TTGTGCCACT GCACTCCAGC 3336  
 CTGGGTGATA GAGTGAGACT CTGTCTCAA AAAAAAAAAA AAAAAAAAAA AAAAACTGA 3396  
 AGGAATTATT CCTCAGGATT TGGGTCTAAT TTGCCCTGAG CACCAACTCC TGAGTTCAAC 3456  
 TACCATGGCT AGACACACCT TAACATTTTC TAGAATCCAC CAGCTTTAGT GGAGTCTGTC 3516  
 TAATCATGAG TATTGGAATA GGATCTGGGG GCAGTGAGGG GGTGGCAGCC ACGTGTGGCA 3576  
 GAGAAAAGCA CACAAGGAAA GAGCACCAG GACTGTCATA TGGAAGAAAG ACAGGACTGC 3636  
 AACTCACCCT TCACAAAATG AGGACCAGAC ACAGCTGATG GTATGAGTTG ATGCAGGTGT 3696  
 GTGGAGCCTC AACATCCTGC TCCCCTCCTA CTACACATGG TTAAGGCCTG TTGCTCTGTC 3756  
 TCCAG GT TCA CAC TCT CTG CAC TAC CTC TTC ATG GGT GCC TCA GAG 3802  
 Arg Ser His Ser Leu His Tyr Leu Phe Met Gly Ala Ser Glu  
 30 35  
 CAG GAC CTT GGT CTT TCC TTG TTT GAA GCT TTG GGC TAC GTG GAT GAC 3850  
 Gln Asp Leu Gly Leu Ser Leu Phe Glu Ala Leu Gly Tyr Val Asp Asp  
 40 45 50 55  
 CAG CTG TTC GTG TTC TAT GAT CAT GAG AGT CGC CGT GTG GAG CCC CGA 3898  
 Gln Leu Phe Val Phe Tyr Asp His Glu Ser Arg Arg Val Glu Pro Arg  
 60 65 70  
 ACT CCA TGG GTT TCC AGT AGA ATT TCA AGC CAG ATG TGG CTG CAG CTG 3946  
 Thr Pro Trp Val Ser Ser Arg Ile Ser Ser Gln Met Trp Leu Gln Leu  
 75 80 85  
 Ser Gln Ser Leu Lys Gly Trp Asp His Met Phe Thr Val Asp Phe Trp  
 90 95 100  
 ACT ATT ATG GAA AAT CAC AAC CAC AGC AAG G GTATGTGGAG AGGGGGCCTC 4045  
 Thr Ile Met Glu Asn His Asn His Ser Lys  
 105 110  
 ACCTTCCTGA GGTGTCAGA GCTTTTCATC TTTTCATGCA TCTTGAAGGA AACAGCTGGA 4105  
 AGTCTGAGGT CTTGTGGGAG CAGGGAAGAG GGAAGGAATT TGCTTCCTGA GATCATTTGG 4165  
 TCCTTGGGGA TGGTGGAAAT AGGGACCTAT TCCTTTGGTT GCAGTTAACA AGGCTGGGGA 4225  
 TTTTCCAG AG TCC CAC ACC CTG CAG GTC ATC CTG GGC TGT GAA ATG 4272  
 Glu Ser His Thr Leu Gln Val Ile Leu Gly Cys Glu Met  
 115 120 125  
 CAA GAA GAC AAC AGT ACC GAG GGC TAC TGG AAG TAC GGG TAT GAT GGG 4320  
 Gln Glu Asp Asn Ser Thr Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly  
 130 135 140  
 CAG GAC CAC CTT GAA TTC TGC CCT GAC ACA CTG GAT TGG AGA GCA GCA 4368

Gln Asp His Leu Glu Phe Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala  
 145 150 155  
 Glu Pro Arg Ala Trp Pro Thr Lys Leu Glu Trp Glu Arg His Lys Ile  
 160 165 170  
 CCG GCC AGG CAG AAC AGG GCC TAC CTG GAG AGG GAC TGC CCT GCA CAG 4464  
 Arg Ala Arg Gln Asn Arg Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln  
 175 180 185 190  
 CTG CAG CAG TTG CTG GAG CTG GGG AGA GGT GTT TTG GAC CAA CAA G 4510  
 Leu Gln Gln Leu Leu Glu Leu Gly Arg Gly Val Leu Asp Gln Gln  
 195 200 205  
 GTATGGTGGG AACACACTTC TGCCCCTATA CTCTAGTGGC AGAGTGGAGG AGGTTGCAGG 4570  
 GCACGGAATC CCTGGTTGGA GTTTCAGAGG TGGCTGAGGC TGTGTGCCTC TCCAAATTCT 4630  
 GGGAAAGGGAC TTTCTCAATC CTAGAGTCTC TACCTTATAA TTGAGATGTA TGAGACAGCC 4690  
 ACAAGTCATG GGTTTAATTT CTTTTCTCCA TGCATATGGC TCAAAGGGAA GTGTCTATGG 4750  
 CCCTTGCTTT TTATTTAACC AATAATCTTT TGTATATTTA TACCTGTTAA AAATTCAGAA 4810  
 ATGTCAAGGC CGGGCAGCGT GGCTCAGCCC TGTAATCCCA GCACTTTGGG AGGCCGAGGC 4870  
 GGGTGGTCAC AAGGTCAGGA GTTTGAGACC AGCCTGACCA ACATGGTGAA ACCCGTCTCT 4930  
 AAAAAAATAC AAAAATTAGC TGGTCACAGT CATGCGCACC TGTAGTCCCA GCTAATTGGA 4990  
 AGGCTGAGGC AGGAGCATCG CTTGAACCTG GGAAGCGGAA GTTGCACTGA GCCAAGATCG 5050  
 CGCCACTGCA CTCCAGCCTA GGCAGCAGAG TGAGACTCCA TCTTAAAAAA AAAAAAAAAA 5110  
 AAAAAAGAG AATTCAGAGA TCTCAGCTAT CATATGAATA CCAGGACAAA ATATCAAGTG 5170  
 AGGCCACTTA TCAGAGTAGA AGAATCCTTT AGGTTAAAG TTTCTTTCAT AGAATATAGC 5230  
 AATAATCACT GAAGCTACCT ATCTTACAAG TCCGCTTCTT ATAACAATGC CTCCTAGGTT 5290  
 GACCCAGGTG AAAGTGACCA TCTGTATTCA ATCATTTTCA ATGCACATAA AGGGCAATTT 5350  
 TATCTATCAG AACAAAGAAC ATGGGTAACA GATATGTATA TTTACATGTG AGGAGAACAA 5410  
 GCTGATCTGA CTGCTCTCCA AGTGACACTG TGTTAGAGTC CAATCTTAGG ACACAAAATG 5470  
 GTGTCTCTCC TGTAGCTTGT TTTTCTCTGA AAAGGGTATT TCCTTCCTCC AACCTATAGA 5530  
 AGGAAGTGAA AGTCCAGTC TTCCTGGCAA GGGTAAACAG ATCCCCTCTC CTCATCCTTC 5590  
 CTCTTTCCTG TCAAG TG CCT CCT TTG GTG AAG GTG ACA CAT CAT GTG ACC 5640  
 Val Pro Pro Leu Val Lys Val Thr His His Val Thr  
 210 215  
 TCT TCA GTG ACC ACT CTA CGG TGT CGG GCC TTG AAC TAC TAC CCC CAG 5688  
 Ser Ser Val Thr Thr Leu Arg Cys Arg Ala Leu Asn Tyr Tyr Pro Gln  
 220 225 230  
 AAC ATC ACC ATG AAG TGG CTG AAG GAT AAG CAG CCA ATG GAT GCC AAG 5736  
 Asn Ile Thr Met Lys Trp Leu Lys Asp Lys Gln Pro Met Asp Ala Lys  
 235 240 245

GAG TTC GAA CCT AAA GAC GTA TTG CCC AAT GGG GAT GGG ACC TAC CAG Glu Phe Glu Pro Lys Asp Val Leu Pro Asn Gly Asp Gly Thr Tyr Gln 250 255 260 265	5784
GGC TGG ATA ACC TTG GCT GTA CCC CCT GGG GAA GAG CAG AGA TAT ACG Gly Trp Ile Thr Leu Ala Val Pro Pro Gly Glu Glu Gln Arg Tyr Thr 270 275 280	5832
TGC CAG GTG GAG CAC CCA GGC CTG GAT CAG CCC CTC ATT GTG ATC TGG G Cys Gln Val Glu His Pro Gly Leu Asp Gln Pro Leu Ile Val Ile Trp 285 290 295	5881
GTATGTGACT GATGAGAGCC AGGAGCTGAG AAAATCTATT GGGGGTTGAG AGGAGTGCCT	5941
GAGGAGGTAA TTATGGCAGT GAGATGAGGA TCTGCTCTTT GTTAGGGGGT GGGCTGAGGG	6001
TGGCAATCAA AGGCTTTAAC TTGCTTTTTC TGTTTTAG AG CCC TCA CCG TCT Glu Pro Ser Pro Ser 300	6053
GGC ACC CTA GTC ATT GGA GTC ATC AGT GGA ATT GCT GTT TTT GTC GTC Gly Thr Leu Val Ile Gly Val Ile Ser Gly Ile Ala Val Phe Val Val 305 310 315	6101
ATC TTG TTC ATT GGA ATT TTG TTC ATA ATA TTA AGG AAG AGG CAG GGT Ile Leu Phe Ile Gly Ile Leu Phe Ile Ile Leu Arg Lys Arg Gln Gly 320 325 330	6149
TCA A GTGAGTAGGA ACAAGGGGGA AGTCTCTTAG TACCTCTGCC CCAGGGCACA Ser 335	6203
GTGGGAAGAG GGGCAGAGGG GATCTGGCAT CCATGGGAAG CATTTTTCTC ATTTATATTC	6263
TTTGGGGACA CCAGCAGCTC CCTGGGAGAC AGAAAATAAT GGTTCCTCCC AGAATGAAAG	6323
TCTCTAATTC AACAAACATC TTCAGAGCAC CTACTATTTT GCAAGAGCTG TTTAAGGTAG	6383
TACAGGGGCT TTGAGGTTGA GAAGTCACTG TGGCTATTCT CAGAACCCAA ATCTGGTAGG	6443
GAATGAABATT GATAGCAAGT AAATGTAGTT AAAGAAGACC CCATGAGGTC CTAAAGCAGG	6503
CAGGAAGCAA ATGCTTAGGG TGTCAAAGGA AAGAATGATC ACATTCAGCT GGGGATCAAG	6563
ATAGCCTTCT GGATCTTGAA GGAGAAGCTG GATTCCATTA GGTGAGGTTG AAGATGATGG	6623
GAGGTCTACA CAGACGGAGC AACCATGCCA AGTAGGAGAG TATAAGGCAT ACTGGGAGAT	6683
TAGAAATAAT TACTGTACCT TAACCCTGAG TTTGCGTAGC TATCACTCAC CAATTATGCA	6743
TTTCTACCCC CTGAACATCT GTGGTGTAGG GAAAAGAGAA TCAGAAAGAA GCCAGCTCAT	6803
ACAGAGTCCA AGGGTCTTTT GGGATATTGG GTTATGATCA CTGGGGTGTC ATTGAAGGAT	6863
CCTAAGAAAG GAGGACCACG ATCTCCCTTA TATGGTGAAT GTGTTGTTAA GAAGTTAGAT	6923
GAGAGGTGAG GAGACCAATT AGAAAGCCAA TAAGCATTTT CAGATGAGAG ATAATGGTTC	6983
TTGAAATCCA ATAGTGCCCA GGTCTAAATT GAGATGGGTG AATGAGGAAA ATAAGGAAGA	7043

GAGAAGAGGC AAGATGGTGC CTAGGTTTGT GATGCCTCTT TCCTGGGTCT CTTGTCTCCA 7103

CAG GA GGA GCC ATG GGG CAC TAC GTC TTA GCT GAA CGT GAG 7144  
Arg Gly Ala Met Gly His Tyr Val Leu Ala Glu Arg Glu  
340 345

TGACACGCAG CCTGCAGACT CACTGTGGGA AGGAGACAAA ACTAGAGACT CAAAGAGGGA 7204

GTGCATTTAT GAGCTCTTCA TGTTCAGGA GAGAGTTGAA CCTAAACATA GAAATTGCCT 7264

GACGAACTCC TTGATTTTAG CCTTCTCTGT TCATTTCTCTC AAAAAAGATTT CCCCATTTAG 7324

GTTTCTGAGT TCCTGCATGC CGGTGATCCC TAGCTGTGAC CTCTCCCCTG GAACTGTCTC 7384

TCATGAACCT CAAGCTGCAT CTAGAGGCTT CCTTCATTTT CTCCGTCACC TCAGAGACAT 7444

ACACCTATGT CATTTTCATTT CCTATTTTGT GAAGAGGACT CCTTAAATTT GGGGGACTTA 7504

CATGATTCAT TTTAACATCT GAGAAAAGCT TTGAACCCTG GGACGTGGCT AGTCATAACC 7564

TTACCAGATT TTTACACATG TATCTATGCA TTTTCTGGAC CCGTTCAACT TTTCTTTTGA 7624

ATCCTCTCTC TGTGTTACCC AGTAACTCAT CTGTCACCAA GCCTTGGGGA TTCTTCCATC 7684

TGATTGTGAT GTGAGTTGCA CAGCTATGAA GGCTGTACAC TGCACGAATG GAAGAGGCAC 7744

CTGTCCCAGA AAAAGCATCA TGGCTATCTG TGGGTAGTAT GATGGGTGTT TTTAGCAGGT 7804

AGGAGGCAAAA TATCTTGAAA GGGGTTGTGA AGAGGTGTTT TTTCTAATTG GCATGAAGGT 7864

GTCATACAGA TTTGCAAAGT TTAATGGTGC CTTCATTTGG GATGCTACTC TAGTATTCCA 7924

GACCTGAAGA ATCACAATAA TTTTCTACCT GGTCTCTCCT TGTTCGATA ATGAAAATTA 7984

TGATAAGGAT GATAAAAGCA CTTACTTCGT GTCCGACTCT TCTGAGCACC TACTTACATG 8044

CATTACTGCA TGCACTTCTT ACAATAATTC TATGAGATAG GTACTATTAT CCCCATTTCT 8104

TTTTTAAATG AAGAAAGTGA AGTAGGCCGG GCACGGTGGC TCACGCCTGT AATCCCAGCA 8164

CTTTGGGAGG CCAABGCGGG TGGATCACGA GGTCAGGAGA TCGAGACCAT CCTGGCTAAC 8224

ATGGTGAAAC CCCATCTCTA ATAAAAATAC AAAAAATTAG CTGGGCGTGG TGGCAGACGC 8284

CTGTAGTCCC AGCTACTCGG AAGGCTGAGG CAGGAGAATG GCATGAACCC AGGAGGCAGA 8344

GCTTGCACTG AGCCGAGTTT GCGCCACTGC ACTCCAGCCT AGGTGACAGA GTGAGACTCC 8404

ATCTCAAAA AATAAAAATA AAAATAAAAA AATGAAAAA AAAAGAAAGT GAAGTATAGA 8464

GATCTCATA GTTTGTCACT GATAGAAACA GGTTTCAAAC TCAGTCAATC TGACCGTTTG 8524

ATACATCTCA GACACCACTA CATTCACTAG TTTAGATGCC TAGAATAAAT AGAGAAGGAA 8584

GGAGATGGCT CTTCTCTTGT CTCATTGTGT TTCTTCTGAG TGAGCTTGAA TCACATGAAG 8644

GGGAACAGCA GAAAACAACC AACTGATCCT CAGCTGTGAT GTTTCCTTTA AAAGTCCCTG 8704

AAGGAAGGTC CTGGAATGTG ACTCCCTTGC TCCTCTGTTG CTCTCTTTGG CATTCAATTC 8764

TTTGGACCCT ACGCAAGGAC TGTAATTGGT GGGGACAGCT AGTGGCCCTG CTGGGCTTCA 8824

CACACGGTGT CCTCCCTAGG CCAGTGCCTC TGGAGTCAGA ACTCTGGTGG TATTTCCCTC	8884
AATGAAGTGG AGTAAGCTCT CTCATTTTGA GATGGTATAA TGGAAGCCAC CAAGTGGCTT	8944
AGAGGATGCC CAGGTCCTTC CATGGAGCCA CTGGGGTTCC GGTGCACATT AAAAAAAAAA	9004
TCTAACCAGG ACATTACAGG ATTGCTAGAT TCTGGGAAAT CAGTTCACCA TGTTCAAAAG	9064
AGTCTTTTTT TTTTTTTTGA GACTCTATTG CCCAGGCTGG AGTGCAATGG CATGATCTCG	9124
GCTCACTGTA ACCTCTGCCT CCCAGGTTCA AGCGATTCTC CTGTCTCAGC CTCCCAAGTA	9184
GCTGGGATTA CAGGCGTGCA CCACCATGCC CGGCTAATTT TTGTATTTTT AGTAGAGACA	9244
GGGTTTCACC ATGTTGGCCA GGCTGGTCTC GAACTCTCCT GACCTCGTGA TCCGCTGCC	9304
TCGGCCTCCC AAAGTGCTGA GATTACAGGT GTGAGCCACC CTGCCAGCC GTCAAAAGAG	9364
TCTTAATATA TATATCCAGA TGGCATGTGT TTACTTTATG TTACTACATG CACTTGGCTG	9424
CATAAATGTG GTACAAGCAT TCTGTCTTGA AGGGCAGGTG CTTCAGGATA CCATATACAG	9484
CTCAGAAGTT TCTTCTTTAG GCATTAAATT TTAGCAAAGA TATCTCATCT CTTCTTTTAA	9544
ACCATTTTCT TTTTTTGTGG TTAGAAAAGT TATGTAGAAA AAAGTAAATG TGATTTACGC	9604
TCATTGTAGA AAAGCTATAA AATGAATACA ATTAAGCTG TTATTTAATT AGCCAGTGAA	9664
AAACTATTAA CAACTTGTCT ATTACCTGTT AGTATTATTG TTGCATTAAA AATGCATATA	9724
CTTTAATAAA TGTATATTGT ATTGTATACT GCATGATTTT ATTGAAGTTC TTGTTTCATCT	9784
TGTGTATATA CTTAATCGCT TTGTCATTTT GGAGACATTT ATTTTGCTTC TAATTTCTTT	9844
ACATTTTGTC TTACGGAATA TTTTCATTCA ACTGTGGTAG CCGAATTAAT CGTGTTCCTT	9904
CACTCTAGGG ACATTGTCGT CTAAGTTGTA AGACATTGGT TATTTTACCA GCAAACCATT	9964
CTGAAAGCAT ATGACAAATT ATTTCTCTCT TAATATCTTA CTATACTGAA AGCAGACTGC	10024
TATAAGGCTT CACTTACTCT TCTACCTCAT AAGGAATATG TTACAATTAA TTTATTAGGT	10084
AAGCATTGTG TTTATATTGG TTTTATTTCA CCTGGGCTGA GATTTCAAGA AACACCCAG	10144
TCTTCACAGT AACACATTTT ACTAACACAT TTAATAAACA TCAGCAACTG TGGCCTGTTA	10204
ATTTTTTTTAA TAGAAATTTT AAGTCCTCAT TTTCTTTTCGG TGTTTTTTAA GCTTAATTTT	10264
TCTGGCTTTA TTCATAAATT CTTAAGGTCA ACTACATTTG AAAAATCAAA GACCTGCATT	10324
TTAAATTCTT ATTCACCTCT GGCAAAACCA TTCACAAACC ATGGTAGTAA AGAGAAGGGT	10384
GACACCTGGT GGCCATAGGT AAATGTACCA CGGTGGTCCG GTGACCAGAG ATGCAGCGCT	10444
GAGGGTTTTT CTGAAGGTAA AGGAATAAAG AATGGGTGGA GGGGCGTGCA CTGGAAATCA	10504
CTTGTAGAGA AAAGCCCCTG AAAATTTGAG AAAACAAACA AGAACTACT TACCAGCTAT	10564
TTGAATTGCT GGAATCACAG GCCATTGCTG AGCTGCCTGA ACTGGGAACA CAACAGAAGG	10624

AAAACAAACC ACTCTGATAA TCATTGAGTC AAGTACAGCA GGTGATTGAG GACTGCTGAG 10684  
 AGGTACAGGC CAAAATTCTT ATGTTGTATT ATAATAATGT CATCTTATAA TACTGTCTAGT 10744  
 ATTTTATAAA ACATTCTTCA CAAACTCACA CACATTTAAA AACAAAACAC TGTCTCTAAA 10804  
 ATCCCCAAAT TTTTCATAAA C 10825

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 348 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Gly Pro Arg Ala Arg Pro Ala Leu Leu Leu Leu Met Leu Leu Gln  
 1 5 10 15  
 Thr Ala Val Leu Gln Gly Arg Leu Leu Arg Ser His Ser Leu His Tyr  
 20 25 30  
 Leu Phe Met Gly Ala Ser Glu Gln Asp Leu Gly Leu Ser Leu Phe Glu  
 35 40 45  
 Ala Leu Gly Tyr Val Asp Asp Gln Leu Phe Val Phe Tyr Asp His Glu  
 50 55 60  
 Ser Arg Arg Val Glu Pro Arg Thr Pro Trp Val Ser Ser Arg Ile Ser  
 65 70 75 80  
 Ser Gln Met Trp Leu Gln Leu Ser Gln Ser Leu Lys Gly Trp Asp His  
 85 90 95  
 Met Phe Thr Val Asp Phe Trp Thr Ile Met Glu Asn His Asn His Ser  
 100 105 110  
 Lys Glu Ser His Thr Leu Gln Val Ile Leu Gly Cys Glu Met Gln Glu  
 115 120 125  
 Asp Asn Ser Thr Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly Gln Asp  
 130 135 140  
 His Leu Glu Phe Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala Glu Pro  
 145 150 155 160  
 Arg Ala Trp Pro Thr Lys Leu Glu Trp Glu Arg His Lys Ile Arg Ala  
 165 170 175  
 Arg Gln Asn Arg Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln Leu Gln  
 180 185 190  
 Gln Leu Leu Glu Leu Gly Arg Gly Val Leu Asp Gln Gln Val Pro Pro  
 195 200 205  
 Leu Val Lys Val Thr His His Val Thr Ser Ser Val Thr Thr Leu Arg  
 210 215 220

```

Cys Arg Ala Leu Asn Tyr Tyr Pro Gln Asn Ile Thr Met Lys Trp Leu
225                      230                      235                      240

Lys Asp Lys Gln Pro Met Asp Ala Lys Glu Phe Glu Pro Lys Asp Val
                      245                      250                      255

                      260                      265                      270

Pro Pro Gly Glu Glu Gln Arg Tyr Thr Cys Gln Val Glu His Pro Gly
                      275                      280                      285

Leu Asp Gln Pro Leu Ile Val Ile Trp Glu Pro Ser Pro Ser Gly Thr
290                      295                      300

Leu Val Ile Gly Val Ile Ser Gly Ile Ala Val Phe Val Val Ile Leu
305                      310                      315                      320

Phe Ile Gly Ile Leu Phe Ile Ile Leu Arg Lys Arg Gln Gly Ser Arg
                      325                      330                      335

Gly Ala Met Gly His Tyr Val Leu Ala Glu Arg Glu
                      340                      345

```

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10825 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: join(361..436, 3762..4025, 4235..4510, 5606..5881, 6040..6153, 7107..7147)
- (D) OTHER INFORMATION: /product= "Hereditary Hemochromatosis (HH) protein containing the 24d1 mutation"
- /note= "Hereditary Hemochromatosis (HH) gene 24d1 allele"

## (ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 140..7319
- (D) OTHER INFORMATION: /note= "start and stop positions for 24d1 allele cDNA (SEQ ID NO:10)"

## (ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 3852..3891
- (D) OTHER INFORMATION: /note= "start and stop positions for genomic sequence surrounding variant for 24d2(C) allele (SEQ ID NO:41)"

## (ix) FEATURE:

- (A) NAME/KEY: -

(B) LOCATION: 5507..6023

(D) OTHER INFORMATION: /note= "start and stop positions for  
genomic sequence surrounding variant  
for 24d1(A) allele (SEQ ID NO:21)"

## (ix) FEATURE:

(A) NAME/KEY: allele

(B) LOCATION: replace(5834, "a")

(D) OTHER INFORMATION: /phenotype= "Hereditary Hemochromatosis  
(HH)"  
/label= 24d1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TCTAAGGTTG AGATAAAATT TTAAATGTA TGATTGAATT TTGAAAATCA TAAATATTTA	60
AATATCTAAA GTTCAGATCA GAACATTGCG AAGCTACTTT CCCCAATCAA CAACACCCCT	120
TCAGGATTTA AAAACCAAGG GGGACACTGG ATCACCTAGT GTTTCACAAG CAGGTACCTT	180
CTGCTGTAGG AGAGAGAGAA CTAAAGTTCT GAAAGACCTG TTGCTTTTCA CCAGGAAGTT	240
TTACTGGGCA TCTCCTGAGC CTAGGCAATA GCTGTAGGGT GACTTCTGGA GCCATCCCCG	300
TTTCCCCGCC CCCCAAAGA AGCGGAGATT TAACGGGGAC GTGCGGCCAG AGCTGGGGAA	360
ATG GGC CCG CGA GCC AGG CCG GCG CTT CTC CTC CTG ATG CTT TTG CAG	408
Met Gly Pro Arg Ala Arg Pro Ala Leu Leu Leu Leu Met Leu Leu Gln	
1 5 10 15	
ACC GCG GTC CTG CAG GGG CGC TTG CTG C GTGAGTCCGA GGGCTGCGGG	456
Thr Ala Val Leu Gln Gly Arg Leu Leu	
20 25	
CGAAGTAGGG GCGCGGCGGG GGTGGAAAAA TCGAACTAG CTTTTCTTT GCGCTTGGGA	516
GTTTGCTAAC TTTGGAGGAC CTGCTCAACC CTATCCGCAA GCCCCTCTCC CTAAGTTCTG	576
CGTCCAGACC CCGTGAGGGA GTGCCTACCA CTGAAGTCA GATAGGGGTC CCTCGCCCCA	636
GGACCTGCCC CCTCCCCCGG CTGTCCCGGC TCTGCGGAGT GACTTTTGGA ACCGCCCCACT	696
CCCTTCCCCC AACTAGAATG CTTTTAAATA AATCTCGTAG TTCCTCACTT GAGCTGAGCT	756
AAGCCTGGGG CTCCTTGAAC CTGGAAGTCG GGTATTATTC CAATGTCAGC TGTGCAGTTT	816
TTTCCCCAGT CATCTCCAAA CAGGAAGTTC TTCCCTGAGT GCTTGCCGAG AAGGCTGAGC	876
AAACCCACAG CAGGATCCGC ACGGGGTTTC CACCTCAGAA CGAATGCGTT GGGCGGTGGG	936
GGCGCGAAG ACTGGCGTTG GGGATCTGAA TTCTTCACCA TTCCACCCAC TTTGGTGAG	996
ACCTGGGGTG GAGGTCTCTA GGGTGGGAGG CTCCTGAGAG AGGCCTACCT CGGGCCTTTC	1056
CCCACTCTTG GCAATTGTTT TTTTGCTGG AAAATTAAGT ATATGTTAGT TTTGAACGTT	1116
TGAAGTGAAC AATTCTCTTT TCGGCTAGGC TTTATTGATT TGCAATGTGC TGTGTAATTA	1176
AGAGGCCTCT CTACAAAGTA CTGATAATGA ACATGTAAGC AATGCACTCA CTTCTAAGTT	1236

ACATTCATAT CTGATCTTAT TTGATTTTCA CTAGGCATAG GGAGGTAGGA GCTAATAATA	1296
CGTTTATTTT ACTAGAAGTT AACTGGAATT CAGATTATAT AACTCTTTTC AGGTTACAAA	1356
GAACATAAAT AATCTGGTTT TCTGATGTTA TTTCAAGTAC TACAGCTGCT TCTAATCTTA	1416
GTTGACAGTG ATTTTGCCCT GTAGTGTAGC ACAGTGTCT GTGGGTCACA CGCCGGCCTC	1476
AGCACAGCAC TTTGAGTTTT GGTACTACGT GTATCCACAT TTTACACATG ACAAGAATGA	1536
GGCATGGCAC GGCCTGCTTC CTGGCAAATT TATTCAATGG TACACTGGGC TTTGGTGGCA	1596
GAGCTCATGT CTCCACTTCA TAGCTATGAT TCTTAAACAT CACACTGCAT TAGAGGTTGA	1656
ATAATAAAT TTCATGTTGA GCAGAAATAT TCATTGTTTA CAAGTGTAAG TGAGTCCCAG	1716
CCATGTGTTG CACTGTTCAA GCCCCAAGGG AGAGAGCAGG GAAACAAGTC TTTACCCTTT	1776
GATATTTTGC ATTCTAGTGG GAGAGATGAC AATAAGCAAA TGAGCAGAAA GATATACAAC	1836
ATCAGGAAAT CATGGGTGTT GTGAGAAGCA GAGAAGTCAG GGCAAGTCAC TCTGGGGCTG	1896
ACACTTGAGC AGAGACATGA AGGAAATAAG AATGATATTG ACTGGGAGCA GTATTTCCCA	1956
GGCAAACTGA GTGGGCCTGG CAAGTTGGAT TAAAAAGCGG GTTTTCTCAG CACTACTCAT	2016
GTGTGTGTGT GTGGGGGGGG GGGGCGGCGT GGGGGTGGGA AGGGGGACTA CCATCTGCAT	2076
GTAGGATGTC TAGCAGTATC CTGTCCTCCC TACTCACTAG GTGCTAGGAG CACTCCCCCA	2136
GTCTTGACAA CCAAAAATGT CTCTAAACTT TGCCACATGT CACCTAGTAG ACAAACCTCT	2196
GGTTAAGAAG CTCGGGTTGA AAAAAATAAA CAAGTAGTGC TGGGGAGTAG AGGCCAAGAA	2256
GTAGGTAATG GGCTCAGAAG AGGAGCCACA AACAAAGGTTG TGCAGGCGCC TGTAGGCTGT	2316
GGTGTGAATT CTAGCCAAGG AGTAACAGTG ATCTGTCACA GGCTTTTAAA AGATTGCTCT	2376
GGCTGCTATG TGGAAGCAG AATGAAGGGA GCAACAGTAA AAGCAGGGAG CCCAGCCAGG	2436
AAGCTGTTAC ACAGTCCAGG CAAGAGGTAG TGGAGTGGGC TGGGTGGGAA CAGAAAAGGG	2496
AGTGACAAAC CATTGTCTCC TGAATATATT CTGAAGGAAG TTGCTGAAGG ATTCTATGTT	2556
GTGTGAGAGA AAGAGAAGAA TTGGCTGGGT GTAGTAGCTC ATGCCAAGGA GGAGGCCAAG	2616
GAGAGCAGAT TCCTGAGCTC AGGAGTTCAA GACCAGCCTG GGCAACACAG CAAAACCCCT	2676
TCTCTACAAA AAATACAAA ATTAGCTGGG TGTGGTGGCA TGCACCTGTG ATCCTAGCTA	2736
CTCGGGAGGC TGAGGTGGAG GGTATTGCTT GAGCCCAGGA AGTTGAGGCT GCAGTGAGCC	2796
ATGACTGTGC CACTGTACTT CAGCCTAGGT GACAGAGCAA GACCCTGTCT CCCCTGACCC	2856
CCTGAAAAG AGAAGAGTTA AAGTTGACTT TGTTCTTTAT TTTAATTTTA TTGGCCTGAG	2916
CAGTGGGGTA ATTGGCAATG CCATTTCTGA GATGGTGAAG GCAGAGGAAA GAGCAGTTTG	2976
GGGTAAATCA AGGATCTGCA TTTGGGACAT GTTAAGTTTG AGATTCCAGT CAGGCTTCCA	3036
AGTGGTGAGG CCACATAGGC AGTTCAGTGT AAGAATTCAG GACCAAGGCT GGGCACGGTG	3096

GCTCACTTCT GTAATCCCAG CACTTTGGTG GCTGAGGCAG GTAGATCATT TGAGGTCAGG 3156  
 AGTTTGAGAC AAGCTTGGCC AACATGGTGA AACCCCATGT CTAATAAAAA TACAAAAATT 3216  
 AGCCTGGTGT GGTGGCGCAC GCCTATAGTC CCAGGTTTTT AGGAGGCTTA GGTAGGAGAA 3276  
 TCCCTTGAAC CCAGGAGGTG CAGGTTGCAG TGAGCTGAGA TTGTGCCACT GCACTCCAGC 3336  
 CTGGGTGATA GAGTGAGACT CTGTCTCAA AAAAAAAAAA AAAAAAAAAA AAAAACTGA 3396  
 AGGAATTATT CCTCAGGATT TGGGTCTAAT TTGCCCTGAG CACCAACTCC TGAGTTCAAC 3456  
 TACCATGGCT AGACACACCT TAACATTTTC TAGAATCCAC CAGCTTTAGT GGAGTCTGTC 3516  
 TAATCATGAG TATTGGAATA GGATCTGGGG GCAGTGAGGG GGTGGCAGCC ACGTGTGGCA 3576  
 GAGAAAAGCA CACAAGGAAA GAGCACCCAG GACTGTCATA TGGAAGAAAG ACAGGACTGC 3636  
 AACTCACCCCT TCACAAAATG AGGACCAGAC ACAGCTGATG GTATGAGTTG ATGCAGGTGT 3696  
 GTGGAGCCTC AACATCCTGC TCCCCTCCTA CTACACATGG TTAAGGCCTG TTGCTCTGTC 3756  
 TCCAG GT TCA CAC TCT CTG CAC TAC CTC TTC ATG GGT GCC TCA GAG 3802  
 Arg Ser His Ser Leu His Tyr Leu Phe Met Gly Ala Ser Glu  
 30 35  
 CAG GAC CTT GGT CTT TCC TTG TTT GAA GCT TTG GGC TAC GTG GAT GAC 3850  
 Gln Asp Leu Gly Leu Ser Leu Phe Glu Ala Leu Gly Tyr Val Asp Asp  
 40 45 50 55  
 CAG CTG TTC GTG TTC TAT GAT CAT GAG AGT CGC CGT GTG GAG CCC CGA 3898  
 Gln Leu Phe Val Phe Tyr Asp His Glu Ser Arg Arg Val Glu Pro Arg  
 60 65 70  
 ACT CCA TGG GTT TCC AGT AGA ATT TCA AGC CAG ATG TGG CTG CAG CTG 3946  
 Thr Pro Trp Val Ser Ser Arg Ile Ser Ser Gln Met Trp Leu Gln Leu  
 75 80 85  
 AGT CAG AGT CTG AAA GGG TGG GAT CAC ATG TTC ACT GTT GAC TTC TGG 3994  
 Ser Gln Ser Leu Lys Gly Trp Asp His Met Phe Thr Val Asp Phe Trp  
 90 95 100  
 ACT ATT ATG GAA AAT CAC AAC CAC AGC AAG G GTATGTGGAG AGGGGGCCTC 4045  
 Thr Ile Met Glu Asn His Asn His Ser Lys  
 105 110  
 ACCTTCCTGA GGTGTGTCAGA GCTTTTCATC TTTTCATGCA TCTTGAAGGA AACAGCTGGA 4105  
 AGTCTGAGGT CTTGTGGGAG CAGGGAAGAG GGAAGGAATT TGCTTCCTGA GATCATTG 4165  
 TCCTTGGGGA TGGTGGAAAT AGGGACCTAT TCCTTTGGTT GCAGTTAACA AGGCTGGGGA 4225  
 TTTTCCAG AG TCC CAC ACC CTG CAG GTC ATC CTG GGC TGT GAA ATG 4272  
 Glu Ser His Thr Leu Gln Val Ile Leu Gly Cys Glu Met  
 115 120 125  
 CAA GAA GAC AAC AGT ACC GAG GGC TAC TGG AAG TAC GGG TAT GAT GGG 4320  
 Gln Glu Asp Asn Ser Thr Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly  
 130 135 140

CAG GAC CAC CTT GAA TTC TGC CCT GAC ACA CTG GAT TGG AGA GCA GCA Gln Asp His Leu Glu Phe Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala 145 150 155	4368
GAA CCC AGG GCC TGG CCC ACC AAG CTG GAG TGG GAA AGG CAC AAG ATT Glu Pro Arg Ala Trp Pro Thr Lys Leu Glu Trp Glu Arg His Lys Ile 160 165 170	4416
CGG GCC AGG CAG AAC AGG GCC TAC CTG GAG AGG GAC TGC CCT GCA CAG Arg Ala Arg Gln Asn Arg Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln 175 - 180 185 190	4464
CTG CAG CAG TTG CTG GAG CTG GGG AGA GGT GTT TTG GAC CAA CAA G Leu Gln Gln Leu Leu Glu Leu Gly Arg Gly Val Leu Asp Gln Gln 195 200 205	4510
GTATGGTGGG AACACACTTC TGCCCCCTATA CTCTAGTGGC AGAGTGGAGG AGGTTGCAGG	4570
GCACGGAATC CCTGGTTGGA GTTTCAGAGG TGGCTGAGGC TGTGTGCCTC TCCAAATTCT	4630
GGGAAGGGAC TTTCTCAATC CTAGAGTCTC TACCTTATAA TTGAGATGTA TGAGACAGCC	4690
ACAAGTCATG GGTTTAATTT CTTTTCTCCA TGCATATGGC TCAAAGGGAA GTGTCTATGG	4750
CCCTTGCTTT TTATTTAACC AATAATCTTT TGTATATTTA TACCTGTAA AAATTCAGAA	4810
ATGTCAAGGC CGGGCACGGT GGCTCACCCC TGTAATCCCA GCACTTTGGG AGGCCGAGGC	4870
GGGTGGTCAC AAGGTCAGGA GTTTGAGACC AGCCTGACCA ACATGGTGAA ACCCGTCTCT	4930
AAAAAATAC AAAAATTAGC TGGTCACAGT CATGCGCACC TGTAGTCCCA GCTAATTGGA	4990
AGGCTGAGGC AGGAGCATCG CTTGAACCTG GGAAGCGGAA GTTGCACTGA GCCAAGATCG	5050
CGCCACTGCA CTCCAGCCTA GGCAGCAGAG TGAGACTCCA TCTTAAAAAA AAAAAAAAAA	5110
AAAAAAGAG AATTCAGAGA TCTCAGCTAT CATATGAATA CCAGGACAAA ATATCAAGTG	5170
AGGCCACTTA TCAGAGTAGA AGAATCCTTT AGGTTAAAAG TTTCTTTCAT AGAACATAGC	5230
AATAATCACT GAAGCTACCT ATCTTACAAG TCCGCTTCTT ATAACAATGC CTCCTAGGTT	5290
GACCCAGGTG AACTGACCA TCTGTATTCA ATCATTTTCA ATGCACATAA AGGGCAATTT	5350
TATCTATCAG AACAAAGAAC ATGGGTAACA GATATGTATA TTTACATGTG AGGAGAACAA	5410
GCTGATCTGA CTGCTCTCCA AGTGACACTG TGTTAGAGTC CAATCTTAGG ACACAAAATG	5470
GTGTCTCTCC TGTAGCTTGT TTTTTTCTGA AAAGGGTATT TCCTTCCTCC AACCTATAGA	5530
AGGAAGTGAA AGTTCCAGTC TTCCTGGCAA GGGTAAACAG ATCCCCTCTC CTCATCCTTC	5590
CTCTTTCCTG TCAAG TG CCT CCT TTG GTG AAG GTG ACA CAT CAT GTG ACC Val Pro Pro Leu Val Lys Val Thr His His Val Thr 210 215	5640
TCT TCA GTG ACC ACT CTA CGG TGT CGG GCC TTG AAC TAC TAC CCC CAG Ser Ser Val Thr Thr Leu Arg Cys Arg Ala Leu Asn Tyr Tyr Pro Gln 220 225 230	5688
AAC ATC ACC ATG AAG TGG CTG AAG GAT AAG CAG CCA ATG GAT GCC AAG	5736

Asn	Ile	Thr	Met	Lys	Trp	Leu	Lys	Asp	Lys	Gln	Pro	Met	Asp	Ala	Lys	
235						240					245					
GAG	TTC	GAA	CCT	AAA	GAC	GTA	TTG	CCC	AAT	GGG	GAT	GGG	ACC	TAC	CAG	5784
Glu	Phe	Glu	Pro	Lys	Asp	Val	Leu	Pro	Asn	Gly	Asp	Gly	Thr	Tyr	Gln	
250					255					260					265	
GGC	TGG	ATA	ACC	TTG	GCT	GTA	CCC	CCT	GGG	GAA	GAG	CAG	AGA	TAT	ACG	5832
Gly	Trp	Ile	Thr		Leu	Ala	Val	Pro	Pro	Gly	Glu	Glu	Gln	Arg	Tyr	
					270					275					280	
TAC	CAG	GTG	GAG	CAC	CCA	GGC	CTG	GAT	CAG	CCC	CTC	ATT	GTG	ATC	TGG	5881
Tyr	Gln	Val	Glu	His	Pro	Gly	Leu	Asp	Gln	Pro	Leu	Ile	Val	Ile	Trp	
			285					290						295		
GTATGTGACT	GATGAGAGCC	AGGAGCTGAG	AAAATCTATT	GGGGGTTGAG	AGGAGTGCCT											5941
GAGGAGGTAA	TTATGGCAGT	GAGATGAGGA	TCTGCTCTTT	GTAGGGGGT	GGGCTGAGGG											6001
TGGCAATCAA	AGGCTTTAAC	TTGCTTTTTC	TGTTTTAG	AG	CCC	TCA	CCG	TCT								6053
									Glu	Pro	Ser	Pro	Ser			
															300	
GGC	ACC	CTA	GTC	ATT	GGA	GTC	ATC	AGT	GGA	ATT	GCT	GTT	TTT	GTC	GTC	6101
Gly	Thr	Leu	Val	Ile	Gly	Val	Ile	Ser	Gly	Ile	Ala	Val	Phe	Val	Val	
			305					310						315		
ATC	TTG	TTC	ATT	GGA	ATT	TTG	TTC	ATA	ATA	TTA	AGG	AAG	AGG	CAG	GGT	6149
Ile	Leu	Phe	Ile	Gly	Ile	Leu	Phe	Ile	Ile	Leu	Arg	Lys	Arg	Gln	Gly	
			320					325						330		
TCA	A	GTGAGTAGGA	ACAAGGGGGA	AGTCTCTTAG	TACCTCTGCC	CCAGGGCACA										6203
Ser																
335																
GTGGGAAGAG	GGGCAGAGGG	GATCTGGCAT	CCATGGGAAG	CATTTTCTC	ATTTATATTC											6263
TTTGGGGACA	CCAGCAGCTC	CCTGGGAGAC	AGAAAATAAT	GTTCTCCCC	AGAATGAAAG											6323
TCTCTAATTC	AACAAACATC	TTCAGAGCAC	CTACTATTTT	GCAAGAGCTG	TTTAAGGTAG											6383
TACAGGGGCT	TTGAGGTTGA	GAAGTCACTG	TGGCTATTCT	CAGAACCCAA	ATCTGGTAGG											6443
GAATGAAATT	GATAGCAAGT	AAATGTAGTT	AAAGAAGACC	CCATGAGGTC	CTAAAGCAGG											6503
CAGGAAGCAA	ATGCTTAGGG	TGTCAAAGGA	AAGAATGATC	ACATTCAGCT	GGGGATCAAG											6563
ATAGCCTTCT	GGATCTTGAA	GGAGAAGCTG	GATTCCATTA	GGTGAGGTTG	AAGATGATGG											6623
GAGGTCTACA	CAGACGGAGC	AACCATGCCA	AGTAGGAGAG	TATAAGGCAT	ACTGGGAGAT											6683
TAGAAATAAT	TACTGTACCT	TAACCCTGAG	TTTGCCTAGC	TATCACTCAC	CAATTATGCA											6743
TTTCTACCCC	CTGAACATCT	GTGGTGTAGG	GAAAAGAGAA	TCAGAAAGAA	GCCAGCTCAT											6803
ACAGAGTCCA	AGGGTCTTTT	GGGATATTGG	GTTATGATCA	CTGGGGTGTC	ATTGAAGGAT											6863
CCTAAGAAAG	GAGGACCACG	ATCTCCCTTA	TATGGTGAAT	GTGTTGTTAA	GAAGTTAGAT											6923
GAGAGGTGAG	GAGACCAGTT	AGAAAGCCAA	TAAGCATTTT	CAGATGAGAG	ATAATGGTTC											6983

TTGAAATCCA ATAGTGCCCA GGTCTAAATT GAGATGGGTG AATGAGGAAA ATAAGGAAGA	7043
GAGAAGAGGC AAGATGGTGC CTAGGTTTGT GATGCCTCTT TCCTGGGTCT CTTGTCTCCA	7103
CAG GA GGA GCC ATG GGG CAC TAC GTC TTA GCT GAA CGT GAG Arg Gly Ala Met Gly His Tyr Val Leu Ala Glu Arg Glu	7144
340 345	
TGACACGCAG CCTGCAGACT CACTGTGGGA AGGAGACAAA ACTAGAGACT CAAAGAGGGA	7204
GTGCATTTAT GAGCTCTTCA TGTTTCAGGA GAGAGTTGAA CCTAAACATA GAAATTGCCT	7264
GACGAACTCC TTGATTTTAG CTTTCTCTGT TCATTTCTCTC AAAAAGATTT CCCCATTTAG	7324
GTTTCTGAGT TCCTGCATGC CGGTGATCCC TAGCTGTGAC CTCTCCCCTG GAACTGTCTC	7384
TCATGAACCT CAAGCTGCAT CTAGAGGCTT CTTTCATTTT CTCCGTCACC TCAGAGACAT	7444
ACACCTATGT CATTTCAATT CCTATTTTTG GAAGAGGACT CCTTAAATTT GGGGGACTTA	7504
CATGATTCAT TTAAACATCT GAGAAAAGCT TTGAACCCTG GGACGTGGCT AGTCATAACC	7564
TTACCAGATT TTTACACATG TATCTATGCA TTTTCTGGAC CCGTTCAACT TTTCTTTTGA	7624
ATCCTCTCTC TGTGTTACCC AGTAACTCAT CTGTCACCAA GCCTTGGGGA TTCTTCCATC	7684
TGATTGTGAT GTGAGTTGCA CAGCTATGAA GGCTGTACAC TGCACGAATG GAAGAGGCAC	7744
CTGTCCCAGA AAAAGCATCA TGGCTATCTG TGGGTAGTAT GATGGGTGTT TTTAGCAGGT	7804
AGGAGGCAAA TATCTTGAAA GGGGTTGTGA AGAGGTGTTT TTTCTAATTG GCATGAAGGT	7864
GTCATACAGA TTTGCAAAGT TTAATGGTGC CTTCAATTGG GATGCTACTC TAGTATTCCA	7924
GACCTGAAGA ATCACAATAA TTTTCTACCT GGTCTCTCCT TGTTCGATA ATGAAAATTA	7984
TGATAAGGAT GATAAAAGCA CTTACTTCGT GTCCGACTCT TCTGAGCACC TACTTACATG	8044
CATTACTGCA TGCACTTCTT ACAATAATTC TATGAGATAG GTACTATTAT CCCCATTCT	8104
TTTTTAAATG AAGAAAGTGA AGTAGGCCGG GCACGGTGGC TCACGCCTGT AATCCCAGCA	8164
CTTTGGGAGG CCAAAGCGGG TGGATCACGA GGTGAGGAGA TCGAGACCAT CCTGGCTAAC	8224
ATGGTGAAAC CCCATCTCTA ATAAAAATAC AAAAAATTAG CTGGGCGTGG TGGCAGACGC	8284
CTGTAGTCCC AGCTACTCGG AAGGCTGAGG CAGGAGAATG GCATGAACCC AGGAGGCAGA	8344
GCTTGCACTG AGCCGAGTTT GCGCCACTGC ACTCCAGCCT AGGTGACAGA GTGAGACTCC	8404
ATCTCAAAAA AATAAAAATA AAAATAAAAA AATGAAAAAA AAAAGAAAGT GAAGTATAGA	8464
GTATCTCATA GTTGTGAGT GATAGAAACA GGTTCCAAAC TCAGTCAATC TGACCGTTTG	8524
ATACATCTCA GACACCACTA CATTCACTAG TTTAGATGCC TAGAATAAAT AGAGAAGGAA	8584
GGAGATGGCT CTTCTCTTGT CTCATTGTGT TTCTTCTGAG TGAGCTTGAA TCACATGAAG	8644
GGGAACAGCA GAAAACAACC AACTGATCCT CAGCTGTCAT GTTTCCTTTA AAAGTCCCTG	8704
AAGGAAGGTC CTGGAATGTG ACTCCCTTGC TCCTCTGTTG CTCTCTTTGG CATTCAATTC	8764

TTTGGACCCT	ACGCAAGGAC	TGTAATTGGT	GGGGACAGCT	AGTGGCCCTG	CTGGGCTTCA	8824
CACACGGTGT	CCTCCCTAGG	CCAGTGCCTC	TGGAGTCAGA	ACTCTGGTGG	TATTTCCCTC	8884
AATGAAGTGG	AGTAAGCTCT	CTCATTTTGA	GATGGTATAA	TGGAAGCCAC	CAAGTGGCTT	8944
AGAGGATGCC	CAGGTCCTTC	CATGGAGCCA	CTGGGGTTCC	GGTGCACATT	AAAAAAAAAA	9004
TCTAACCAGG	ACATTCAGGA	ATTGCTAGAT	TCTGGGAAAT	CAGTTCACCA	TGTTCAAAAG	9064
AGTCTTTTTT	TTTTTTTTGA	GACTCTATTG	CCCAGGCTGG	AGTGCAATGG	CATGATCTCG	9124
GCTCACTGTA	ACCTCTGCCT	CCCAGGTTCA	AGCGATTCTC	CTGTCTCAGC	CTCCCAAGTA	9184
GCTGGGATTA	CAGGCGTGCA	CCACCATGCC	CGGCTAATTT	TTGTATTTTT	AGTAGAGACA	9244
GGGTTTCACC	ATGTTGGCCA	GGCTGGTCTC	GAACTCTCCT	GACCTCGTGA	TCCGCCTGCC	9304
TCGGCCTCCC	AAAGTGCTGA	GATTACAGGT	GTGAGCCACC	CTGCCCAGCC	GTCAAAAGAG	9364
TCTTAATATA	TATATCCAGA	TGGCATGTGT	TTACTTTATG	TTACTACATG	CACTTGGCTG	9424
CATAAATGTG	GTACAAGCAT	TCTGTCTTGA	AGGGCAGGTG	CTTCAGGATA	CCATATACAG	9484
CTCAGAAAGT	TCTTCTTTAG	GCATTAAATT	TTAGCAAAGA	TATCTCATCT	CTTCTTTTAA	9544
ACCATTTTCT	TTTTTTGTGG	TTAGAAAAGT	TATGTAGAAA	AAAGTAAATG	TGATTTACGC	9604
TCATTGTAGA	AAAGCTATAA	AATGAATACA	ATTAAAGCTG	TTATTTAATT	AGCCAGTGAA	9664
AAACTATTAA	CAACTTGTCT	ATTACCTGTT	AGTATTATTG	TTGCATTAAA	AATGCATATA	9724
CTTTAATAAA	TGTATATTGT	ATTGTATACT	GCATGATTTT	ATTGAAGTTC	TTGTTTCATCT	9784
TGTGTATATA	CTTAATCGCT	TTGTCATTTT	GGAGACATTT	ATTTTGCTTC	TAATTTCTTT	9844
ACATTTTGTG	TTACGGAATA	TTTTCATTC	ACTGTGGTAG	CCGAATTAAT	CGTGTCTTCT	9904
CACTCTAGGG	ACATTGTGCT	CTAAGTTGTA	AGACATTGGT	TATTTTACCA	GCAAACCATT	9964
CTGAAAGCAT	ATGACAAATT	ATTTCTCTCT	TAATATCTTA	CTATACTGAA	AGCAGACTGC	10024
TATAAGGCTT	CACTTACTCT	TCTACCTCAT	AAGGAATATG	TTACAATTAA	TTTATTAGGT	10084
AAGCATTTGT	TTTATATTGG	TTTTATTTC	CCTGGGCTGA	GATTTCAAGA	AACACCCCAG	10144
TCTTCACAGT	AACACATTTT	ACTAACACAT	TTACTAAACA	TCAGCAACTG	TGGCCTGTTA	10204
ATTTTTTTAA	TAGAAATTTT	AAGTCCTCAT	TTTCTTTCGG	TGTTTTTTAA	GCTTAATTTT	10264
TCTGGCTTTA	TTCATAAATT	CTTAAGGTCA	ACTACATTTG	AAAAATCAAA	GACCTGCATT	10324
TTAAATTCTT	ATTCACCTCT	GGCAAAACCA	TTACAAACC	ATGGTAGTAA	AGAGAAGGGT	10384
GACACCTGGT	GGCCATAGGT	AAATGTACCA	CGGTGGTCCG	GTGACCAGAG	ATGCAGCGCT	10444
GAGGGTTTTT	CTGAAGGTAA	AGGAATAAAG	AATGGGTGGA	GGGGCGTGCA	CTGGAAATCA	10504
CTTGTAAGAGA	AAAGCCCCTG	AAAATTTGAG	AAAACAAACA	AGAACTACT	TACCAGCTAT	10564

```

TTGAATTGCT GGAATCACAG GCCATTGCTG AGCTGCCTGA ACTGGGAACA CAACAGAAGG 10624
AAAACAAACC ACTCTGATAA TCATTGAGTC AAGTACAGCA GGTGATTGAG GACTGCTGAG 10684
AGGTACAGGC CAAAATTCTT ATGTTGTATT ATAATAATGT CATCTTATAA TACTGTCAGT 10744
ATTTTATAAA ACATTCTTCA CAAACTCACA CACATTTAAA AACAAAACAC TGTCTCTAAA 10804
ATCCCCAAAT TTTTCATAAA C 10825

```

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 348 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

Met Gly Pro Arg Ala Arg Pro Ala Leu Leu Leu Leu Met Leu Leu Gln
 1             5             10             15
Thr Ala Val Leu Gln Gly Arg Leu Leu Arg Ser His Ser Leu His Tyr
          20             25             30
Leu Phe Met Gly Ala Ser Glu Gln Asp Leu Gly Leu Ser Leu Phe Glu
          35             40             45
Ala Leu Gly Tyr Val Asp Asp Gln Leu Phe Val Phe Tyr Asp His Glu
          50             55             60
Ser Arg Arg Val Glu Pro Arg Thr Pro Trp Val Ser Ser Arg Ile Ser
          65             70             75             80
Ser Gln Met Trp Leu Gln Leu Ser Gln Ser Leu Lys Gly Trp Asp His
          85             90             95
Met Phe Thr Val Asp Phe Trp Thr Ile Met Glu Asn His Asn His Ser
          100            105            110
Lys Glu Ser His Thr Leu Gln Val Ile Leu Gly Cys Glu Met Gln Glu
          115            120            125
Asp Asn Ser Thr Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly Gln Asp
          130            135            140
His Leu Glu Phe Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala Glu Pro
          145            150            155            160
Arg Ala Trp Pro Thr Lys Leu Glu Trp Glu Arg His Lys Ile Arg Ala
          165            170            175
Arg Gln Asn Arg Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln Leu Gln
          180            185            190
Gln Leu Leu Glu Leu Gly Arg Gly Val Leu Asp Gln Gln Val Pro Pro
          195            200            205

```

```

Leu Val Lys Val Thr His His Val Thr Ser Ser Val Thr Thr Leu Arg
210                      215                      220

Cys Arg Ala Leu Asn Tyr Tyr Pro Gln Asn Ile Thr Met Lys Trp Leu
225                      230                      235                      240

Lys Asp Lys Gln Pro Met Asp Ala Lys Glu Phe Glu Pro Lys Asp Val
                245                      250                      255

Leu Pro Asn Gly Asp Gly Thr Tyr Gln Gly Trp Ile Thr Leu Ala Val
                260                      265                      270

Pro Pro Gly Glu Glu Gln Arg Tyr Thr Tyr Gln Val Glu His Pro Gly
275                      280                      285

Leu Asp Gln Pro Leu Ile Val Ile Trp Glu Pro Ser Pro Ser Gly Thr
290                      295                      300

Leu Val Ile Gly Val Ile Ser Gly Ile Ala Val Phe Val Val Ile Leu
305                      310                      315                      320

Phe Ile Gly Ile Leu Phe Ile Ile Leu Arg Lys Arg Gln Gly Ser Arg
                325                      330                      335

Gly Ala Met Gly His Tyr Val Leu Ala Glu Arg Glu
                340                      345

```

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10825 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: join(361..436, 3762..4025, 4235..4510, 5606..5881, 6040..6153, 7107..7147)
- (D) OTHER INFORMATION: /product= "Hereditary Hemochromatosis (HH) protein containing the 24d2 mutation"  
/note= "Hereditary Hemochromatosis (HH) gene 24d2 allele"

## (ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 140..7319
- (D) OTHER INFORMATION: /note= "start and stop positions for 24d2 allele cDNA (SEQ ID NO:11)"

## (ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 3852..3891
- (D) OTHER INFORMATION: /note= "start and stop positions for genomic sequence surrounding variant for 24d2(G) allele (SEQ ID NO:42)"

## (ix) FEATURE:

(A) NAME/KEY: -

(B) LOCATION: 5507..6023

(D) OTHER INFORMATION: /note= "start and stop positions for  
genomic sequence surrounding variant  
for 24d1(G) allele (SEQ ID NO:20)"

## (ix) FEATURE:

(A) NAME/KEY: allele

(B) LOCATION: replace(3872, "g")

(D) OTHER INFORMATION: /phenotype= "Hereditary Hemochromatosis  
(HH)"  
/label= 24d2

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TCTAAGGTTG AGATAAAATT TTTAAATGTA TGATTGAATT TTGAAAATCA TAAATATTTA	60
AATATCTAAA GTTCAGATCA GAACATTGCG AAGCTACTTT CCCCAATCAA CAACACCCCT	120
TCAGGATTTA AAAACCAAGG GGGACACTGG ATCACCTAGT GTTTCACAAG CAGGTACCTT	180
CTGCTGTAGG AGAGAGAGAA CTAAAGTTCT GAAAGACCTG TTGCTTTTCA CCAGGAAGTT	240
TTACTGGGCA TCTCCTGAGC CTAGGCAATA GCTGTAGGGT GACTTCTGGA GCCATCCCCG	300
TTTCCCCGCC CCCCAAAGA AGCGGAGATT TAACGGGGAC GTGCGGCCAG AGCTGGGGAA	360
ATG GGC CCG CGA GCC AGG CCG GCG CTT CTC CTC CTG ATG CTT TTG CAG	408
Met Gly Pro Arg Ala Arg Pro Ala Leu Leu Leu Leu Met Leu Leu Gln	
1 5 10 15	
ACC GCG GTC CTG CAG GGG CGC TTG CTG C GTGAGTCCGA GGGCTGCGGG	456
Thr Ala Val Leu Gln Gly Arg Leu Leu	
20 25	
CGAACTAGGG GCGCGGCGGG GGTGGAAAAA TCGAACTAG CTTTTCTTT GCGCTTGGA	516
GTTTGCTAAC TTTGGAGGAC CTGCTCAACC CTATCCGCAA GCCCTCTCC CTACTTTCTG	576
CGTCCAGACC CCGTGAGGGA GTGCCTACCA CTGAACTGCA GATAGGGGTC CCTCGCCCCA	636
GGACCTGCCC CCTCCCCCGG CTGTCCCGGC TCTGCGGAGT GACTTTTGGA ACCGCCCACT	696
CCCTTCCCCC AACTAGAATG CTTTTAAATA AATCTCGTAG TTCCTCACTT GAGCTGAGCT	756
AAGCCTGGGG CTCCTTGAAC CTGGAACTCG GGTTTATTTT CAATGTCAGC TGTGCAGTTT	816
TTTCCCCAGT CATCTCCAAA CAGGAAGTTC TTCCCTGAGT GCTTGCCGAG AAGGCTGAGC	876
AAACCCACAG CAGGATCCGC ACGGGGTTTC CACCTCAGAA CGAATGCGTT GGGCGGTGGG	936
GGCGCGAAAG AGTGGCGTTG GGGATCTGAA TTCTTCACCA TTCCACCCAC TTTTGGTGAG	996
ACCTGGGGTG GAGGTCTCTA GGGTGGGAGG CTCCTGAGAG AGGCCTACCT CGGGCCTTTC	1056
CCCCTCTTG GCAATTGTTT TTTTGCCTGG AAAATTAAGT ATATGTTAGT TTTGAACGTT	1116
TGAACTGAAC AATTCTCTTT TCGGCTAGGC TTTATTGATT TGCAATGTGC TGTGTAATTA	1176

AGAGGCCTCT CTACAAAGTA CTGATAATGA ACATGTAAGC AATGCACTCA CTTCTAAGTT	1236
ACATTCATAT CTGATCTTAT TTGATTTTCA CTAGGCATAG GGAGGTAGGA GCTAATAATA	1296
CGTTTATTTT ACTAGAAGTT AACTGGAATT CAGATTATAT AACTCTTTTC AGGTTACAAA	1356
GAACATAAAT AATCTGGTTT TCTGATGTTA TTTCAAGTAC TACAGCTGCT TCTAATCTTA	1416
GTTGACAGTG ATTTTGCCCT GTAGTGTAGC ACAGTGTTCT GTGGGTCACA CGCCGGCCTC	1476
AGCACAGCAC TTTGAGTTTT GGTACTACGT GTATCCACAT TTTACACATG ACAAGAATGA	1536
GGCATGGCAC GGCCTGCTTC CTGGCAAATT TATTCAATGG TACACTGGGC TTTGGTGGCA	1596
GAGCTCATGT CTCCACTTCA TAGCTATGAT TCTTAAACAT CACACTGCAT TAGAGGTTGA	1656
ATAATAAAAT TTCATGTTGA GCAGAAATAT TCATTGTTTA CAAGTGTAAG TGAGTCCCAG	1716
CCATGTGTTG CACTGTTCAA GCCCCAAGGG AGAGAGCAGG GAAACAAGTC TTTACCCTTT	1776
GATATTTTGC ATTCTAGTGG GAGAGATGAC AATAAGCAAA TGAGCAGAAA GATATACAAC	1836
ATCAGGAAAT CATGGGTGTT GTGAGAAGCA GAGAAGTCAG GGCAAGTCAC TCTGGGGCTG	1896
ACACTTGAGC AGAGACATGA AGGAAATAAG AATGATATTG ACTGGGAGCA GTATTTCCCA	1956
GGCAAAGTGA GTGGGCCTGG CAAGTTGGAT TAAAAAGCGG GTTTTCTCAG CACTACTCAT	2016
GTGTGTGTGT GTGGGGGGGG GGGGCGGCGT GGGGGTGGGA AGGGGGACTA CCATCTGCAT	2076
GTAGGATGTC TAGCAGTATC CTGTCCTCCC TACTCACTAG GTGCTAGGAG CACTCCCCCA	2136
GTCTTGACAA CCAAAAATGT CTCTAAACTT TGCCACATGT CACCTAGTAG ACAAACTCCT	2196
GGTTAAGAAG CTCGGGTTGA AAAAAATAAA CAAGTAGTGC TGGGAGTAG AGGCCAAGAA	2256
GTAGGTAATG GGCTCAGAAG AGGAGCCACA AACAAGGTTG TGCAGGCGCC TGTAGGCTGT	2316
GGTGTGAATT CTAGCCAAGG AGTAACAGTG ATCTGTCACA GGCTTTTAAA AGATTGCTCT	2376
GGCTGCTATG TGGAAAGCAG AATGAAGGGA GCAACAGTAA AAGCAGGGAG CCCAGCCAGG	2436
AAGCTGTTAC ACAGTCCAGG CAAGAGGTAG TGGAGTGGGC TGGGTGGGAA CAGAAAAGGG	2496
AGTGACAAAC CATTGTCTCC TGAATATATT CTGAAGGAAG TTGCTGAAGG ATTCTATGTT	2556
GTGTGAGAGA AAGAGAAGAA TTGGCTGGGT GTAGTAGCTC ATGCCAAGGA GGAGGCCAAG	2616
GAGAGCAGAT TCCTGAGCTC AGGAGTTCAA GACCAGCCTG GGCAACACAG CAAAACCCCT	2676
TCTCTACAAA AAATACAAA ATTAGCTGGG TGTGGTGGCA TGCACCTGTG ATCCTAGCTA	2736
CTCGGGAGGC TGAGGTGGAG GGTATTGCTT GAGCCCAGGA AGTTGAGGCT GCAGTGAGCC	2796
ATGACTGTGC CACTGTACTT CAGCCTAGGT GACAGAGCAA GACCCTGTCT CCCCTGACCC	2856
CCTGAAAAAG AGAAGAGTTA AAGTTGACTT TGTTCTTTAT TTAATTTTA TTGGCCTGAG	2916
CAGTGGGGTA ATTGGCAATG CCATTTCTGA GATGGTGAAG GCAGAGGAAA GAGCAGTTTG	2976

GGGTAAATCA AGGATCTGCA TTTGGGACAT GTTAAGTTTG AGATTCCAGT CAGGCTTCCA 3036  
 AGTGGTGAGG CCACATAGGC AGTTCAGTGT AAGAATTCAG GACCAAGGCT GGGCACGGTG 3096  
 GCTCACTTCT GTAATCCCAG CACTTTGGTG GCTGAGGCAG GTAGATCATT TGAGGTCAGG 3156  
 AGTTTGAGAC AAGCTTGGCC AACATGGTGA AACCCCATGT CTACTAAAAA TACAAAAATT 3216  
 AGCCTGGTGT GGTGGCGCAC GCCTATAGTC CCAGGTTTTTC AGGAGGCTTA GGTAGGAGAA 3276  
 TCCCTTGAAC CCAGGAGGTG CAGGTTGCAG TGAGCTGAGA TTGTGCCACT GCACTCCAGC 3336  
 CTGGGTGATA GAGTGAGACT CTGTCTCAA AAAAAAAAAA AAAAAAACTGA 3396  
 AGGAATTATT CCTCAGGATT TGGGTCTAAT TTGCCCTGAG CACCAACTCC TGAGTTCAAC 3456  
 TACCATGGCT AGACACACCT TAACATTTTC TAGAATCCAC CAGCTTTAGT GGAGTCTGTC 3516  
 TAATCATGAG TATTGGAATA GGATCTGGGG GCAGTGAGGG GGTGGCAGCC ACGTGTGGCA 3576  
 GAGAAAAGCA CACAAGGAAA GAGCACCCAG GACTGTCATA TGGAAGAAAG ACAGGACTGC 3636  
 AACTCACCTC TCACAAAATG AGGACCAGAC ACAGCTGATG GTATGAGTTG ATGCAGGTGT 3696  
 GTGGAGCCTC AACATCCTGC TCCCCTCCTA CTACACATGG TTAAGGCCTG TTGCTCTGTC 3756  
 TCCAG GT TCA CAC TCT CTG CAC TAC CTC TTC ATG GGT GCC TCA GAG 3802  
 Arg Ser His Ser Leu His Tyr Leu Phe Met Gly Ala Ser Glu  
 30 35  
 CAG GAC CTT GGT CTT TCC TTG TTT GAA GCT TTG GGC TAC GTG GAT GAC 3850  
 Gln Asp Leu Gly Leu Ser Leu Phe Glu Ala Leu Gly Tyr Val Asp Asp  
 40 45 50 55  
 CAG CTG TTC GTG TTC TAT GAT GAT GAG AGT CGC CGT GTG GAG CCC CGA 3898  
 Gln Leu Phe Val Phe Tyr Asp Asp Glu Ser Arg Arg Val Glu Pro Arg  
 60 65 70  
 ACT CCA TGG GTT TCC AGT AGA ATT TCA AGC CAG ATG TGG CTG CAG CTG 3946  
 Thr Pro Trp Val Ser Ser Arg Ile Ser Ser Gln Met Trp Leu Gln Leu  
 75 80 85  
 AGT CAG AGT CTG AAA GGG TGG GAT CAC ATG TTC ACT GTT GAC TTC TGG 3994  
 Ser Gln Ser Leu Lys Gly Trp Asp His Met Phe Thr Val Asp Phe Trp  
 90 95 100  
 ACT ATT ATG GAA AAT CAC AAC CAC AGC AAG G GTATGTGGAG AGGGGGCCTC 4045  
 Thr Ile Met Glu Asn His Asn His Ser Lys  
 105 110  
 ACCTTCCTGA GGTGTGCAGA GCTTTTCATC TTTTCATGCA TCTTGAAGGA AACAGCTGGA 4105  
 AGTCTGAGGT CTTGTGGGAG CAGGGAAGAG GGAAGGAATT TGCTTCCTGA GATCATTG 4165  
 TCCTTGGGGA TGGTGGAAT AGGGACCTAT TCCTTTGGTT GCAGTTAACA AGGCTGGGGA 4225  
 TTTTCCAG AG TCC CAC ACC CTG CAG GTC ATC CTG GGC TGT GAA ATG 4272  
 Glu Ser His Thr Leu Gln Val Ile Leu Gly Cys Glu Met  
 115 120 125  
 CAA GAA GAC AAC AGT ACC GAG GGC TAC TGG AAG TAC GGG TAT GAT GGG 4320

Gln	Glu	Asp	Asn	Ser	Thr	Glu	Gly	Tyr	Trp	Lys	Tyr	Gly	Tyr	Asp	Gly		
			130					135					140				
CAG	GAC	CAC	CTT	GAA	TTC	TGC	CCT	GAC	ACA	CTG	GAT	TGG	AGA	GCA	GCA		4368
Gln	Asp	His	Leu	Glu	Phe	Cys	Pro	Asp	Thr	Leu	Asp	Trp	Arg	Ala	Ala		
		145					150					155					
GAA	CCC	AGG	GCC	TGG	CCC	ACC	AAG	CTG	GAG	TGG	GAA	AGG	CAC	AAG	ATT		4416
Glu	Pro	Arg	Ala	Trp	Pro	Thr	Lys	Leu	Glu	Trp	Glu	Arg	His	Lys	Ile		
	160					165					170						
CGG	GCC	AGG	CAG	AAC	AGG	GCC	TAC	CTG	GAG	AGG	GAC	TGC	CCT	GCA	CAG		4464
Arg	Ala	Arg	Gln	Asn	Arg	Ala	Tyr	Leu	Glu	Arg	Asp	Cys	Pro	Ala	Gln		
175					180					185					190		
CTG	CAG	CAG	TTG	CTG	GAG	CTG	GGG	AGA	GGT	GTT	TTG	GAC	CAA	CAA	G		4510
Leu	Gln	Gln	Leu	Leu	Glu	Leu	Gly	Arg	Gly	Val	Leu	Asp	Gln	Gln			
			195					200					205				
GTATGGTGGG	AACACACTTC	TGCCCCCTATA	CTCTAGTGGC	AGAGTGGAGG	AGGTTGCAGG												4570
GCACGGAATC	CCTGGTTGGA	GTTTCAGAGG	TGGCTGAGGC	TGTGTGCCTC	TCCAAATTCT												4630
GGGAAGGGAC	TTTCTCAATC	CTAGAGTCTC	TACCTTATAA	TTGAGATGTA	TGAGACAGCC												4690
ACAAGTCATG	GGTTTAATTT	CTTTTCTCCA	TGCATATGGC	TCAAAGGGAA	GTGTCTATGG												4750
CCCTTGCTTT	TTATTTAACC	AATAATCTTT	TGTATATTTA	TACCTGTTAA	AAATTCAGAA												4810
ATGTCAAGGC	CGGGCACGGT	GGCTCACCCC	TGTAATCCCA	GCACTTTGGG	AGGCCGAGGC												4870
GGGTGGTCAC	AAGGTCAGGA	GTTTGAGACC	AGCCTGACCA	ACATGGTGAA	ACCCGTCTCT												4930
AAAAAAATAC	AAAAATTAGC	TGGTCACAGT	CATGCGCACC	TGTAGTCCCA	GCTAATTGGA												4990
AGGCTGAGGC	AGGAGCATCG	CTTGAACCTG	GGAAGCGGAA	GTTGCACTGA	GCCAAGATCG												5050
CGCCACTGCA	CTCCAGCCTA	GGCAGCAGAG	TGAGACTCCA	TCTTAAAAAA	AAAAAAAATA												5110
AAAAAAGAG	AATTCAGAGA	TCTCAGCTAT	CATATGAATA	CCAGGACAAA	ATATCAAGTG												5170
AGGCCACTTA	TCAGAGTAGA	AGAATCCTTT	AGGTTAAAG	TTTCTTTCAT	AGAACATAGC												5230
AATAATCACT	GAAGCTACCT	ATCTTACAAG	TCCGCTTCTT	ATAACAATGC	CTCCTAGGTT												5290
GACCCAGGTG	AAACTGACCA	TCTGTATTCA	ATCATTTTCA	ATGCACATAA	AGGGCAATTT												5350
TATCTATCAG	AACAAAGAAC	ATGGGTAACA	GATATGTATA	TTTACATGTG	AGGAGAACAA												5410
GCTGATCTGA	CTGCTCTCCA	AGTGACACTG	TGTTAGAGTC	CAATCTTAGG	ACACAAAATG												5470
GTGTCTCTCC	TGTAGCTTGT	TTTTTTCTGA	AAAGGGTATT	TCCTTCCTCC	AACCTATAGA												5530
AGGAAGTGAA	AGTTCCAGTC	TTCCTGGCAA	GGGTAAACAG	ATCCCCTCTC	CTCATCCTTC												5590
CTCTTTCCTG	TCAAG	TG	CCT	CCT	TTG	GTG	AAG	GTG	ACA	CAT	CAT	GTG	ACC				5640
		Val	Pro	Pro	Leu	Val	Lys	Val	Thr	His	His	Val	Thr				
						210						215					
TCT	TCA	GTG	ACC	ACT	CTA	CGG	TGT	CGG	GCC	TTG	AAC	TAC	TAC	CCC	CAG		5688
Ser	Ser	Val	Thr	Thr	Leu	Arg	Cys	Arg	Ala	Leu	Asn	Tyr	Tyr	Pro	Gln		

220	225	230	
AAC ATC ACC ATG AAG TGG CTG AAG GAT AAG CAG CCA ATG GAT GCC AAG Asn Ile Thr Met Lys Trp Leu Lys Asp Lys Gln Pro Met Asp Ala Lys 235 240 245			5736
GAG TTC GAA CCT AAA GAC GTA TTG CCC AAT GGG GAT GGG ACC TAC CAG Glu Phe Glu Pro Lys Asp Val Leu Pro Asn Gly Asp Gly Thr Tyr Gln			5784
GGC TGG ATA ACC TTG GCT GTA CCC CCT GGG GAA GAG CAG AGA TAT ACG Gly Trp Ile Thr Leu Ala Val Pro Pro Gly Glu Glu Gln Arg Tyr Thr 270 275 280			5832
TGC CAG GTG GAG CAC CCA GGC CTG GAT CAG CCC CTC ATT GTG ATC TGG G Cys Gln Val Glu His Pro Gly Leu Asp Gln Pro Leu Ile Val Ile Trp 285 290 295			5881
GTATGTGACT GATGAGAGCC AGGAGCTGAG AAAATCTATT GGGGTTGAG AGGAGTGCCT			5941
GAGGAGGTAA TTATGGCAGT GAGATGAGGA TCTGCTCTTT GTTAGGGGGT GGGCTGAGGG			6001
TGGCAATCAA AGGCTTTAAC TTGCTTTTTC TGTTTTAG AG CCC TCA CCG TCT Glu Pro Ser Pro Ser 300			6053
GGC ACC CTA GTC ATT GGA GTC ATC AGT GGA ATT GCT GTT TTT GTC GTC Gly Thr Leu Val Ile Gly Val Ile Ser Gly Ile Ala Val Phe Val Val 305 310 315			6101
ATC TTG TTC ATT GGA ATT TTG TTC ATA ATA TTA AGG AAG AGG CAG GGT Ile Leu Phe Ile Gly Ile Leu Phe Ile Ile Leu Arg Lys Arg Gln Gly 320 325 330			6149
TCA A GTGAGTAGGA ACAAGGGGGA AGTCTCTTAG TACCTCTGCC CCAGGGCACA Ser 335			6203
GTGGGAAGAG GGGCAGAGGG GATCTGGCAT CCATGGGAAG CATTTTTCTC ATTTATATTC			6263
TTTGGGGACA CCAGCAGCTC CCTGGGAGAC AGAAAATAAT GGTTCCTCCCC AGAATGAAAG			6323
TCTCTAATTC AACAAACATC TTCAGAGCAC CTACTATTTT GCAAGAGCTG TTAAAGGTAG			6383
TACAGGGGCT TTGAGGTTGA GAAGTCACTG TGGCTATTCT CAGAACCCAA ATCTGGTAGG			6443
GAATGAAATT GATAGCAAGT AAATGTAGTT AAAGAAGACC CCATGAGGTC CTAAAGCAGG			6503
CAGGAAGCAA ATGCTTAGGG TGTCAAAGGA AAGAATGATC ACATTGAGCT GGGGATCAAG			6563
ATAGCCTTCT GGATCTTGAA GGAGAAGCTG GATTCCATTA GGTGAGGTTG AAGATGATGG			6623
GAGGTCTACA CAGACGGAGC AACCATGCCA AGTAGGAGAG TATAAGGCAT ACTGGGAGAT			6683
TAGAAATAAT TACTGTACCT TAACCCTGAG TTTGCGTAGC TATCACTCAC CAATTATGCA			6743
TTTCTACCCC CTGAACATCT GTGGTGTAGG GAAAAGAGAA TCAGAAAGAA GCCAGCTCAT			6803
ACAGAGTCCA AGGGTCTTTT GGGATATTGG GTTATGATCA CTGGGGTGTC ATTGAAGGAT			6863
CCTAAGAAAG GAGGACCACG ATCTCCCTTA TATGGTGAAT GTGTTGTAA GAAGTTAGAT			6923

GAGAGGTGAG GAGACCAGTT AGAAAGCCAA TAAGCATTTT CAGATGAGAG ATAATGGTTC	6983
TTGAAATCCA ATAGTGCCCA GGTCTAAATT GAGATGGGTG AATGAGGAAA ATAAGGAAGA	7043
GAGAAGAGGC AAGATGGTGC CTAGGTTTGT GATGCCTCTT TCCTGGGTCT CTTGTCTCCA	7103
CAG GA GGA GCC ATG GGG CAC TAC GTC TTA GCT GAA CGT GAG Arg Gly Ala Met Gly His Tyr Val Leu Ala Glu Arg Glu 340 345	7144
TGACACGCAG CCTGCAGACT CACTGTGGGA AGGAGACAAA ACTAGAGACT CAAAGAGGGA	7204
GTGCATTTAT GAGCTCTTCA TGTTTCAGGA GAGAGTTGAA CCTAAACATA GAAATTGCCT	7264
GACGAACGCC TTGATTTTAG CCTTCTCTGT TCATTTCTCT CAAAAGATTT CCCCATTTAG	7324
GTTTCTGAGT TCCTGCATGC CGGTGATCCC TAGCTGTGAC CTCTCCCCTG GAACTGTCTC	7384
TCATGAACCT CAAGCTGCAT CTAGAGGCTT CCTTCATTTT CTCCGTCACC TCAGAGACAT	7444
ACACCTATGT CATTTTATTT CCTATTTTTG GAAGAGGACT CCTTAAATTT GGGGGACTTA	7504
CATGATTCAT TTTAACATCT GAGAAAAGCT TTGAACCCCTG GGACGTGGCT AGTCATAACC	7564
TTACCAGATT TTTACACATG TATCTATGCA TTTTCTGGAC CCGTTCAACT TTTCTTTGA	7624
ATCCTCTCTC TGTGTTACCC AGTAACTCAT CTGTCACCAA GCCTTGGGGA TTCTTCCATC	7684
TGATTGTGAT GTGAGTTGCA CAGCTATGAA GGCTGTACAC TGCACGAATG GAAGAGGCAC	7744
CTGTCCCAGA AAAAGCATCA TGGCTATCTG TGGGTAGTAT GATGGGTGTT TTTAGCAGGT	7804
AGGAGGCCAAA TATCTTGAAG GGGGTGTGTA AGAGGTGTTT TTTCTAATTG GCATGAAGGT	7864
GTCATACAGA TTTGCAAAGT TTAATGGTGC CTTTATTTGG GATGCTACTC TAGTATTCCA	7924
GACCTGAAGA ATCACAATAA TTTTCTACCT GGTCTCTCCT TGTTCTGATA ATGAAAATTA	7984
TGATAAGGAT GATAAAAGCA CTTACTTCGT GTCCGACTCT TCTGAGCACC TACTTACATG	8044
CATTACTGCA TGCACTTCTT ACAATAATTC TATGAGATAG GTACTATTAT CCCCATTCTT	8104
TTTTTAAATG AAGAAAGTGA AGTAGGCCGG GCACGGTGGC TCACGCCTGT AATCCCAGCA	8164
CTTTGGGAGG CCAAAGCGGG TGGATCACGA GGTCAGGAGA TCGAGACCAT CCTGGCTAAC	8224
ATGGTGAAAC CCCATCTCTA ATAAAAATAC AAAAAATTAG CTGGGCGTGG TGGCAGACGC	8284
CTGTAGTCCC AGCTACTCGG AAGGCTGAGG CAGGAGAATG GCATGAACCC AGGAGGCAGA	8344
GCTTGCAGTG AGCCGAGTTT GCGCCACTGC ACTCCAGCCT AGGTGACAGA GTGAGACTCC	8404
ATCTCAAAAA AATAAAAAATA AAAATAAAAA AATGAAAAAA AAAAGAAAGT GAAGTATAGA	8464
GTATCTCATA GTTTGTCAGT GATAGAAACA GGTTCAAAC TCAGTCAATC TGACCGTTTG	8524
ATACATCTCA GACACCACTA CATTCACTAG TTTAGATGCC TAGAATAAAT AGAGAAGGAA	8584
GGAGATGGCT CTTCTCTTGT CTCATTGTGT TTCTTCTGAG TGAGCTTGAA TCACATGAAG	8644
GGGAACAGCA GAAAACAACC AACTGATCCT CAGCTGTCAT GTTTCCTTTA AAAGTCCCTG	8704

AAGGAAGGTC CTGGAATGTG ACTCCCTTGC TCCTCTGTTG CTCTCTTTGG CATTCAATTC	8764
TTTGGACCCT ACGCAAGGAC TGTAATTGGT GGGGACAGCT AGTGGCCCTG CTGGGCTTCA	8824
CACACGGTGT CCTCCCTAGG CCAGTGCCTC TGGAGTCAGA ACTCTGGTGG TATTTCCCTC	8884
AATGAAGTGG AGTAAGCTCT CTCATTTTGA GATGGTATAA TGGAAGCCAC CAAGTGGCTT	8944
AGAGGATGCC CAGGTCCTTC CATGGAGCCA CTGGGGTTCC GGTGCACATT AAAAAAAAAA	9004
TCTAACCAGG ACATTCAGGA ATTGCTAGAT TCTGGGAAAT CAGTTCACCA TGTTCAAAAG	9064
AGTCTTTTTT TTTTTTTTGA GACTCTATTG CCCAGGCTGG AGTGCAATGG CATGATCTCG	9124
GCTCACTGTA ACCTCTGCCT CCCAGGTTCA AGCGATTCTC CTGTCTCAGC CTCCCAAGTA	9184
GCTGGGATTA CAGGCGTGCA CCACCATGCC CGGCTAATTT TTGTATTTTT AGTAGAGACA	9244
GGGTTTCACC ATGTTGGCCA GGCTGGTCTC GAACTCTCCT GACCTCGTGA TCCGCCTGCC	9304
TCGGCCTCCC AAAGTGCTGA GATTACAGGT GTGAGCCACC CTGCCCAGCC GTCAAAAGAG	9364
TCTTAATATA TATATCCAGA TGGCATGTGT TTACTTTATG TTACTACATG CACTTGGCTG	9424
CATAAATGTG GTACAAGCAT TCTGTCTTGA AGGGCAGGTG CTTCAGGATA CCATATACAG	9484
CTCAGAAGTT TCTTCTTTAG GCATTAAATT TTAGCAAAGA TATCTCATCT CTTCTTTTAA	9544
ACCATTTTCT TTTTTTGTGG TTAGAAAAGT TATGTAGAAA AAAGTAAATG TGATTTACGC	9604
TCATTGTAGA AAAGCTATAA AATGAATACA ATTAAAGCTG TTATTTAATT AGCCAGTGAA	9664
AAACTATTAA CAACTTGTCT ATTACCTGTT AGTATTATTG TTGCATTAAA AATGCATATA	9724
TGTGTATATA CTTAATCGCT TTGTCATTTT GGAGACATTT ATTTTGCTTC TAATTTCTTT	9844
ACATTTTGTC TTACGGAATA TTTTCATTCA ACTGTGGTAG CCGAATTAAT CGTGTTCCTT	9904
CACTCTAGGG ACATTGTGCT CTAAGTTGTA AGACATTGGT TATTTTACCA GCAAACCATT	9964
CTGAAAGCAT ATGACAAATT ATTTCTCTCT TAATATCTTA CTATACTGAA AGCAGACTGC	10024
TATAAGGCTT CACTTACTCT TCTACCTCAT AAGGAATATG TTACAATTAA TTTATTAGGT	10084
AAGCATTGTG TTTATATTGG TTTTATTTCA CCTGGGCTGA GATTTCAAGA AACACCCAG	10144
TCTTCACAGT AACACATTTT ACTAACACAT TACTAAACA TCAGCAACTG TGGCCTGTTA	10204
ATTTTTTTAA TAGAAATTTT AAGTCCTCAT TTTCTTTCGG TGTTTTTTAA GCTTAATTTT	10264
TCTGGCTTTA TTCATAAATT CTTAAGGTCA ACTACATTTG AAAAATCAAA GACCTGCATT	10324
TTAAATTCTT ATTCACCTCT GGCAAAACCA TTCACAAACC ATGGTAGTAA AGAGAAGGGT	10384
GACACCTGGT GGCCATAGGT AAATGTACCA CGGTGGTCCG GTGACCAGAG ATGCAGCGCT	10444
GAGGGTTTTT CTGAAGGTAA AGGAATAAAG AATGGGTGGA GGGGCGTGCA CTGGAAATCA	10504
CTTGTAGAGA AAAGCCCCTG AAAATTTGAG AAAACAAACA AGAACTACT TACCAGCTAT	10564

```

TTGAATTGCT GGAATCACAG GCCATTGCTG AGCTGCCTGA ACTGGGAACA CAACAGAAGG 10624
AAAACAAACC ACTCTGATAA TCATTGAGTC AAGTACAGCA GGTGATTGAG GACTGCTGAG 10684
AGGTACAGGC CAAAATTCTT ATGTTGTATT ATAATAATGT CATCTTATAA TACTGTCACT 10744
ATTTTATAAA ACATTCTTCA CAAACTCACA CACATTTAAA AACAAAACAC TGTCTCTAAA 10804
ATCCCCAAAT TTTTCATAAA C 10825

```

## (2) INFORMATION FOR SEQ ID NO:6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 348 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```

Met Gly Pro Arg Ala Arg Pro Ala Leu Leu Leu Leu Met Leu Leu Gln
 1             5             10             15
Thr Ala Val Leu Gln Gly Arg Leu Leu Arg Ser His Ser Leu His Tyr
      20             25             30
Leu Phe Met Gly Ala Ser Glu Gln Asp Leu Gly Leu Ser Leu Phe Glu
      35             40             45
Ala Leu Gly Tyr Val Asp Asp Gln Leu Phe Val Phe Tyr Asp Asp Glu
      50             55             60
Ser Arg Arg Val Glu Pro Arg Thr Pro Trp Val Ser Ser Arg Ile Ser
      65             70             75             80
Ser Gln Met Trp Leu Gln Leu Ser Gln Ser Leu Lys Gly Trp Asp His
      85             90             95
Met Phe Thr Val Asp Phe Trp Thr Ile Met Glu Asn His Asn His Ser
      100            105            110
Lys Glu Ser His Thr Leu Gln Val Ile Leu Gly Cys Glu Met Gln Glu
      115            120            125
Asp Asn Ser Thr Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly Gln Asp
      130            135            140
His Leu Glu Phe Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala Glu Pro
      145            150            155            160
Arg Ala Trp Pro Thr Lys Leu Glu Trp Glu Arg His Lys Ile Arg Ala
      165            170            175
Arg Gln Asn Arg Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln Leu Gln
      180            185            190
Gln Leu Leu Glu Leu Gly Arg Gly Val Leu Asp Gln Gln Val Pro Pro
      195            200            205

```

```

Leu Val Lys Val Thr His His Val Thr Ser Ser Val Thr Thr Leu Arg
 210                      215                      220

Cys Arg Ala Leu Asn Tyr Tyr Pro Gln Asn Ile Thr Met Lys Trp Leu
 225                      230                      235                      240

Lys Asp Lys Gln Pro Met Asp Ala Lys Glu Phe Glu Pro Lys Asp Val
                      245                      250                      255

Leu Pro Asn Gly Asp Gly Thr Tyr Gln Gly Trp Ile Thr Leu Ala Val
                      260                      265                      270

Pro Pro Gly Glu Glu Gln Arg Tyr Thr Cys Gln Val Glu His Pro Gly
                      275                      280                      285

Leu Asp Gln Pro Leu Ile Val Ile Trp Glu Pro Ser Pro Ser Gly Thr
                      290                      295                      300

Leu Val Ile Gly Val Ile Ser Gly Ile Ala Val Phe Val Val Ile Leu
 305                      310                      315                      320

Phe Ile Gly Ile Leu Phe Ile Ile Leu Arg Lys Arg Gln Gly Ser Arg
                      325                      330                      335

Gly Ala Met Gly His Tyr Val Leu Ala Glu Arg Glu
                      340                      345

```

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10825 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: join(361..436, 3762..4025, 4235..4510, 5606..5881, 6040..6153, 7107..7147)
- (D) OTHER INFORMATION: /product= "Hereditary Hemochromatosis (HH) protein containing both the 24d1 and 24d2 mutations"  
/note= "Hereditary Hemochromatosis (HH) gene containing a combination of both 24d1 and 24d2 alleles"

## (ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 140..7319
- (D) OTHER INFORMATION: /note= "start and stop positions for cDNA containing a combination of both 24d1 and 24d2 alleles (SEQ ID NO:12)"

## (ix) FEATURE:

- (A) NAME/KEY: -
- (B) LOCATION: 3852..3891

(D) OTHER INFORMATION: /note= "start and stop positions for genomic sequence surrounding variant for 24d2(G) allele (SEQ ID NO:42)"

(ix) FEATURE:

(A) NAME/KEY: -

(B) LOCATION: 5507..6023

(D) OTHER INFORMATION: /note= "start and stop positions for genomic sequence surrounding variant for 24d1(A) allele (SEQ ID NO:21)"

(ix) FEATURE:

(A) NAME/KEY: allele

(B) LOCATION: replace(3872, "g")

(D) OTHER INFORMATION: /phenotype= "Hereditary Hemochromatosis (HH)"  
/label= 24d2

(ix) FEATURE:

(A) NAME/KEY: allele

(B) LOCATION: replace(5834, "a")

(D) OTHER INFORMATION: /phenotype= "Hereditary Hemochromatosis (HH)"  
/label= 24d1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
TCTAAGGTTG AGATAAAATT TTTAAATGTA TGATTGAATT TTGAAAATCA TAAATATTTA      60
AATATCTAAA GTTCAGATCA GAACATTGCG AAGCTACTTT CCCCAATCAA CAACACCCCT      120
TCAGGATTTA AAAACCAAGG GGGACACTGG ATCACCTAGT GTTTCACAAG CAGGTACCTT      180
CTGCTGTAGG AGAGAGAGAA CTAAAGTTCT GAAAGACCTG TTGCTTTTCA CCAGGAAGTT      240
TTACTGGGCA TCTCCTGAGC CTAGGCAATA GCTGTAGGGT GACTTCTGGA GCCATCCCCG      300
TTTCCCCGCC CCCCAAAGA AGCGGAGATT TAACGGGGAC GTGCGGCCAG AGCTGGGGAA      360
ATG GGC CCG CGA GCC AGG CCG GCG CTT CTC CTC CTG ATG CTT TTG CAG      408
Met Gly Pro Arg Ala Arg Pro Ala Leu Leu Leu Leu Met Leu Leu Gln
   1             5             10             15

ACC GCG GTC CTG CAG GGG CGC TTG CTG C GTGAGTCCGA GGGCTGCGGG      456
Thr Ala Val Leu Gln Gly Arg Leu Leu
   20             25

CGAACTAGGG GCGCGGCGGG GGTGGAAGAA TCGAACTAG CTTTTCTTT GCGCTTGGGA      516
GTTTGCTAAC TTTGGAGGAC CTGCTCAACC CTATCCGCAA GCCCCTCTCC CTACTTTCTG      576
CGTCCAGACC CCGTGAGGGA GTGCCTACCA CTGAACTGCA GATAGGGGTC CCTCGCCCCA      636
GGACCTGCCC CCTCCCCCGG CTGTCCCGGC TCTGCGGAGT GACTTTTGGA ACCGCCCCACT      696
CCCTTCCCCC AACTAGAATG CTTTTAAATA AATCTCGTAG TTCCTCACTT GAGCTGAGCT      756
AAGCCTGGGG CTCCTTGAAC CTGGAACTCG GGTATTATTC CAATGTCAGC TGTGCAGTTT      816
TTTCCCAGT CATCTCCAAA CAGGAAGTTC TTCCCTGAGT GCTTGCCGAG AAGGCTGAGC      876
```

AAACCCACAG CAGGATCCGC ACGGGGTTTC CACCTCAGAA CGAATGCGTT GGGCGGTGGG	936
GGCGCGAAAG AGTGGCGTTG GGGATCTGAA TTCTTCACCA TTCCACCCAC TTTTGGTGAG	996
ACCTGGGGTG GAGGTCTCTA GGGTGGGAGG CTCCTGAGAG AGGCCTACCT CGGGCCTTTC	1056
CCCCTCTTG GCAATTGTTC TTTTGCCCTGG AAAATTAAGT ATATGTTAGT TTTGAACGTT	1116
TGAACTGAAC AATTCTCTTT TCGGCTAGGC TTTATTGATT TGCAATGTGC TGTGTAATTA	1176
AGAGGCCTCT CTACAAAGTA CTGATAATGA ACATGTAAGC AATGCACTCA CTTCTAAGTT	1236
ACATTCATAT CTGATCTTAT TTGATTTTCA CTAGGCATAG GGAGGTAGGA GCTAATAATA	1296
CGTTTATTTT ACTAGAAGTT AACTGGAATT CAGATTATAT AACTCTTTTC AGGTTACAAA	1356
GAACATAAAT AATCTGGTTT TCTGATGTTA TTTCAAGTAC TACAGCTGCT TCTAATCTTA	1416
GTTGACAGTG ATTTTGCCCT GTAGTGTAGC ACAGTGTTCT GTGGGTCACA CGCCGGCCTC	1476
AGCACAGCAC TTTGAGTTT GGTACTACGT GTATCCACAT TTTACACATG ACAAGAATGA	1536
GGCATGGCAC GGCCTGCTTC CTGGCAAATT TATTCAATGG TACTCTGGGC TTTGGTGGCA	1596
GAGCTCATGT CTCCACTTCA TAGCTATGAT TCTTAAACAT CACACTGCAT TAGAGGTTGA	1656
ATAATAAAAT TTCATGTTGA GCAGAAATAT TCATTGTTTA CAAGTGTAAG TGAGTCCCAG	1716
CCATGTGTTG CACTGTTCAA GCCCCAAGGG AGAGAGCAGG GAAACAAGTC TTTACCCCTT	1776
GATATTTTGC ATTCTAGTGG GAGAGATGAC AATAAGCAAA TGAGCAGAAA GATATACAAC	1836
ATCAGGAAAT CATGGGTGTT GTGAGAAGCA GAGAAGTCAG GGCAAGTCAC TCTGGGGCTG	1896
ACACTTGAGC AGAGACATGA AGGAAATAAG AATGATATTG ACTGGGAGCA GTATTTCCCA	1956
GGCAAAGTGA GTGGGCCTGG CAAGTTGGAT TAAAAAGCGG GTTTTCTCAG CACTACTCAT	2016
GTGTGTGTGT GTGGGGGGGG GGGGCGGCGT GGGGGTGGGA AGGGGGACTA CCATCTGCAT	2076
GTAGGATGTC TAGCAGTATC CTGTCCTCCC TACTCACTAG GTGCTAGGAG CACTCCCCCA	2136
GTCTTGACAA CAAAAAATGT CTCTAAACTT TGCCACATGT CACCTAGTAG ACAAACCTCT	2196
GGTTAAGAAG CTCGGGTTGA AAAAAATAAA CAAGTAGTGC TGGGGAGTAG AGGCCAAGAA	2256
GTAGGTAATG GGCTCAGAAG AGGAGCCACA AACAAAGTTG TGCAGGCGCC TGTAGGCTGT	2316
GGCTGCTATG TGGAAAGCAG AATGAAGGGA GCAACAGTAA AAGCAGGGAG CCCAGCCAGG	2436
AAGCTGTTAC ACAGTCCAGG CAAGAGGTAG TGGACTGGGC TGGGTGGGAA CAGAAAAGGG	2496
AGTGACAAAC CATTGTCTCC TGAATATATT CTGAAGGAAG TTGCTGAAGG ATTCTATGTT	2556
GTGTGAGAGA AAGAGAAGAA TTGGCTGGGT GTAGTAGCTC ATGCCAAGGA GGAGGCCAAG	2616
GAGAGCAGAT TCCTGAGCTC AGGAGTTCAA GACCAGCCTG GGCAACACAG CAAAACCCCT	2676
TCTCTACAAA AAATACAAA ATTAGCTGGG TGTGGTGGCA TGCACCTGTG ATCCTAGCTA	2736

CTCGGGAGGC TGAGGTGGAG GGTATTGCTT GAGCCCAGGA AGTTGAGGCT GCAGTGAGCC	2796
ATGACTGTGC CACTGTACTT CAGCCTAGGT GACAGAGCAA GACCCTGTCT CCCCTGACCC	2856
CCTGAAAAAG AGAAGAGTTA AAGTTGACTT TGTTCTTTAT TTTAATTTTA TTGGCCTGAG	2916
CAGTGGGGTA ATTGGCAATG CCATTTCTGA GATGGTGAAG GCAGAGGAAA GAGCAGTTTG	2976
GGGTAAATCA AGGATCTGCA TTTGGGACAT GTTAAGTTTG AGATTCCAGT CAGGCTTCCA	3036
AGTGGTGAGG CCACATAGGC AGTTCAGTGT AAGAATTCAG GACCAAGGCT GGGCACGGTG	3096
GCTCACTTCT GTAATCCCAG CACTTTGGTG GCTGAGGCAG GTAGATCATT TGAGGTCAGG	3156
AGTTTGAGAC AAGCTTGGCC AACATGGTGA AACCCCATGT CTACTAAAAA TACAAAAATT	3216
AGCCTGGTGT GGTGGCGCAC GCCTATAGTC CCAGGTTTTT AGGAGGCTTA GGTAGGAGAA	3276
TCCCTTGAAC CCAGGAGGTG CAGGTTGCAG TGAGCTGAGA TTGTGCCACT GCACTCCAGC	3336
CTGGGTGATA GAGTGAGACT CTGTCTCAA AAAAAAAAAA AAAAAAAAAA AAAAAACTGA	3396
AGGAATTATT CCTCAGGATT TGGGTCTAAT TTGCCCTGAG CACCAACTCC TGAGTTCAAC	3456
TACCATGGCT AGACACACCT TAACATTTTC TAGAATCCAC CAGCTTTAGT GGAGTCTGTC	3516
TAATCATGAG TATTGGAATA GGATCTGGGG GCAGTGAGGG GGTGGCAGCC ACGTGTGGCA	3576
GAGAAAAGCA CACAAGGAAA GAGCACCCAG GACTGTCATA TGGAAGAAAG ACAGGACTGC	3636
AACTCACCCCT TCACAAAATG AGGACCAGAC ACAGCTGATG GTATGAGTTG ATGCAGGTGT	3696
GTGGAGCCTC AACATCCTGC TCCCCTCCTA CTACACATGG TTAAGGCCTG TTGCTCTGTC	3756
TCCAG GT TCA CAC TCT CTG CAC TAC CTC TTC ATG GGT GCC TCA GAG	3802
Arg Ser His Ser Leu His Tyr Leu Phe Met Gly Ala Ser Glu	
30 35	
CAG GAC CTT GGT CTT TCC TTG TTT GAA GCT TTG GGC TAC GTG GAT GAC	3850
Gln Asp Leu Gly Leu Ser Leu Phe Glu Ala Leu Gly Tyr Val Asp Asp	
40 45 50 55	
CAG CTG TTC GTG TTC TAT GAT GAT GAG AGT CGC CGT GTG GAG CCC CGA	3898
Gln Leu Phe Val Phe Tyr Asp Asp Glu Ser Arg Arg Val Glu Pro Arg	
60 65 70	
ACT CCA TGG GTT TCC AGT AGA ATT TCA AGC CAG ATG TGG CTG CAG CTG	3946
Thr Pro Trp Val Ser Ser Arg Ile Ser Ser Gln Met Trp Leu Gln Leu	
75 80 85	
AGT CAG AGT CTG AAA GGG TGG GAT CAC ATG TTC ACT GTT GAC TTC TGG	3994
Ser Gln Ser Leu Lys Gly Trp Asp His Met Phe Thr Val Asp Phe Trp	
90 95 100	
ACT ATT ATG GAA AAT CAC AAC CAC AGC AAG G GTATGTGCAG AGGGGGCCTC	4045
Thr Ile Met Glu Asn His Asn His Ser Lys	
105 110	
ACCTTCCTGA GGTGTGAGA GCTTTTCATC TTTTCATGCA TCTTGAAGGA AACAGCTGGA	4105
AGTCTGAGGT CTTGTGGGAG CAGGGAAGAG GGAAGGAATT TGCTTCCTGA GATCATTTGG	4165

TCCTTGGGGA TGGTGGAAAT AGGGACCTAT TCCTTTGGTT GCAGTTAACA AGGCTGGGGA	4225
TTTTTCCAG AG TCC CAC ACC CTG CAG GTC ATC CTG GGC TGT GAA ATG Glu Ser His Thr Leu Gln Val Ile Leu Gly Cys Glu Met 115 120 125	4272
CAA GAA GAC AAC AGT ACC GAG GGC TAC TGG AAG TAC GGG TAT GAT GGG Gln Glu Asp Asn Ser Thr Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly 130 135 140	4320
CAG GAC CAC CTT GAA TTC TGC CCT GAC ACA CTG GAT TGG AGA GCA GCA Gln Asp His Leu Glu Phe Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala 145 150 155	4368
GAA CCC AGG GCC TGG CCC ACC AAG CTG GAG TGG GAA AGG CAC AAG ATT Glu Pro Arg Ala Trp Pro Thr Lys Leu Glu Trp Glu Arg His Lys Ile 160 165 170	4416
CGG GCC AGG CAG AAC AGG GCC TAC CTG GAG AGG GAC TGC CCT GCA CAG Arg Ala Arg Gln Asn Arg Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln 175 180 185 190	4464
CTG CAG CAG TTG CTG GAG CTG GGG AGA GGT GTT TTG GAC CAA CAA G Leu Gln Gln Leu Leu Glu Leu Gly Arg Gly Val Leu Asp Gln Gln 195 200 205	4510
GTATGGTGGA AACACACTTC TGCCCCCTATA CTCTAGTGGC AGAGTGGAGG AGGTTGCAGG	4570
GCACGGAATC CCTGGTTGGA GTTTCAGAGG TGGCTGAGGC TGTGTGCCTC TCCAAATTCT	4630
GGGAAGGGAC TTTCTCAATC CTAGAGTCTC TACCTTATAA TTGAGATGTA TGAGACAGCC	4690
ACAAGTCATG GGTTTAATTT CTTTTCTCCA TGCATATGGC TCAAAGGGAA GTGTCTATGG	4750
CCCTTGCTTT TTATTTAACC AATAATCTTT TGTATATTTA TACCTGTTAA AAATTCAGAA	4810
ATGTCAAGGC CGGGCACGGT GGCTCACCCC TGTAATCCCA GCACTTTGGG AGGCCGAGGC	4870
GGGTGGTCAC AAGGTCAGGA GTTTGAGACC AGCCTGACCA ACATGGTGAA ACCCGTCTCT	4930
AAAAAATAC AAAAATTAGC TGGTCACAGT CATGCGCACC TGTAGTCCCA GCTAATTGGA	4990
AGGCTGAGGC AGGAGCATCG CTTGAACCTG GGAAGCGGAA GTTGCACTGA GCCAAGATCG	5050
CGCCACTGCA CTCCAGCCTA GGCAGCAGAG TGAGACTCCA TCTTAAAAAA AAAAAAAAAA	5110
AAAAAAGAG AATTCAGAGA TCTCAGCTAT CATATGAATA CCAGGACAAA ATATCAAGTG	5170
AGGCCACTTA TCAGAGTAGA AGAATCCTTT AGGTTAAAAG TTTCTTTCAT AGAACATAGC	5230
AATAATCACT GAAGCTACCT ATCTTACAAG TCCGCTTCTT ATAACAATGC CTCCTAGGTT	5290
GACCCAGGTG AACTGACCA TCTGTATTCA ATCATTTTCA ATGCACATAA AGGGCAATTT	5350
TATCTATCAG AACAAAGAAC ATGGGTAACA GATATGTATA TTTACATGTG AGGAGAACAA	5410
GCTGATCTGA CTGCTCTCCA AGTGACACTG TGTTAGAGTC CAATCTTAGG ACACAAAATG	5470
GTGTCTCTCC TGTAGCTTGT TTTTTTCTGA AAAGGGTATT TCCTTCCTCC AACCTATAGA	5530

AGGAAGTGAA AGTTCCAGTC TTCCTGGCAA GGGTAAACAG ATCCCCTCTC CTCATCCTTC 5590

CTCTTTCCCTG TCAAG TG CCT CCT TTG GTG AAG GTG ACA CAT CAT GTG ACC 5640  
Val Pro Pro Leu Val Lys Val Thr His His Val Thr  
210 215

TCT TCA GTG ACC ACT CTA CGG TGT CGG GCC TTG AAC TAC TAC CCC CAG 5688  
Ser Ser Val Thr Thr Leu Arg Cys Arg Ala Leu Asn Tyr Tyr Pro Gln  
220 225 230

AAC ATC ACC ATG AAG TGG CTG AAG GAT AAG CAG CCA ATG GAT GCC AAG 5736  
Asn Ile Thr Met Lys Trp Leu Lys Asp Lys Gln Pro Met Asp Ala Lys  
235 240 245

GAG TTC GAA CCT AAA GAC GTA TTG CCC AAT GGG GAT GGG ACC TAC CAG 5784  
Glu Phe Glu Pro Lys Asp Val Leu Pro Asn Gly Asp Gly Thr Tyr Gln  
250 255 260 265

GGC TGG ATA ACC TTG GCT GTA CCC CCT GGG GAA GAG CAG AGA TAT ACG 5832  
Gly Trp Ile Thr Leu Ala Val Pro Pro Gly Glu Glu Gln Arg Tyr Thr  
270 275 280

TAC CAG GTG GAG CAC CCA GGC CTG GAT CAG CCC CTC ATT GTG ATC TGG G 5881  
Tyr Gln Val Glu His Pro Gly Leu Asp Gln Pro Leu Ile Val Ile Trp  
285 290 295

GTATGTGACT GATGAGAGCC AGGAGCTGAG AAAATCTATT GGGGGTTGAG AGGAGTGCCT 5941

GAGGAGGTAA TTATGGCAGT GAGATGAGGA TCTGCTCTTT GTTAGGGGGT GGGCTGAGGG 6001

TGGCAATCAA AGGCTTTAAC TTGCTTTTTC TGTTTTAG AG CCC TCA CCG TCT 6053  
Glu Pro Ser Pro Ser  
300

GGC ACC CTA GTC ATT GGA GTC ATC AGT GGA ATT GCT GTT TTT GTC GTC 6101  
Gly Thr Leu Val Ile Gly Val Ile Ser Gly Ile Ala Val Phe Val Val  
305 310 315

ATC TTG TTC ATT GGA ATT TTG TTC ATA ATA TTA AGG AAG AGG CAG GGT 6149  
Ile Leu Phe Ile Gly Ile Leu Phe Ile Ile Leu Arg Lys Arg Gln Gly  
320 325 330

TCA A GTGAGTAGGA ACAAGGGGGA AGTCTCTTAG TACCTCTGCC CCAGGGCACA 6203  
Ser  
335

GTGGGAAGAG GGGCAGAGGG GATCTGGCAT CCATGGGAAG CATTTTCTC ATTTATATTC 6263

TTTGGGGACA CCAGCAGCTC CCTGGGAGAC AGAAAATAAT GGTTCCTCCC AGAATGAAAG 6323

TCTCTAATTC AACAAACATC TTCAGAGCAC CTACTATTTT GCAAGAGCTG TTTAAGGTAG 6383

TACAGGGGCT TTGAGGTTGA GAAGTCACTG TGGCTATTCT CAGAACCCAA ATCTGGTAGG 6443

GAATGAAATT GATAGCAAGT AAATGTAGTT AAAGAAGACC CCATGAGGTC CTAAAGCAGG 6503

CAGGAAGCAA ATGCTTAGGG TGTCAAAGGA AAGAATGATC ACATTCAGCT GGGGATCAAG 6563

ATAGCCTTCT GGATCTTGAA GGAGAAGCTG GATTCCATTA GGTGAGGTTG AAGATGATGG 6623

GAGGTCTACA CAGACGGAGC AACCATGCCA AGTAGGAGAG TATAAGGCAT ACTGGGAGAT 6683

TAGAAATAAT TACTGTACCT TAACCCTGAG TTTGCGTAGC TATCACTCAC CAATTATGCA	6743
TTTCTACCCC CTGAACATCT GTGGTGTAGG GAAAAGAGAA TCAGAAAGAA GCCAGCTCAT	6803
ACAGAGTCCA AGGGTCTTTT GGGATATTGG GTTATGATCA CTGGGGTGTG ATTGAAGGAT	6863
CCTAAGAAAG GAGGACCACG ATCTCCCTTA TATGGTGAAT GTGTTGTTAA GAAGTTAGAT	6923
GAGAGGTGAG GAGACCAGTT AGAAAGCCAA TAAGCATTTT CAGATGAGAG ATAATGGTTC	6983
TTGAAATCCA ATAGTGCCCA GGTCTAAATT GAGATGGGTG AATGAGGAAA ATAAGGAAGA	7043
GAGAAGAGGC AAGATGGTGC CTAGGTTTGT GATGCCTCTT TCCTGGGTCT CTTGTCTCCA	7103
CAG GA GGA GCC ATG GGG CAC TAC GTC TTA GCT GAA CGT GAG Arg Gly Ala Met Gly His Tyr Val Leu Ala Glu Arg Glu	7144
340 345	
TGACACGCAG CCTGCAGACT CACTGTGGGA AGGAGACAAA ACTAGAGACT CAAAGAGGGA	7204
GTGCATTTAT GAGCTCTTCA TGTTTCAGGA GAGAGTTGAA CCTAAACATA GAAATTGCCT	7264
GACGAACTCC TTGATTTTAG CCTTCTCTGT TCATTTCTCTC AAAAAGATTT CCCCATTTAG	7324
GTTTCTGAGT TCCTGCATGC CGGTGATCCC TAGCTGTGAC CTCTCCCCTG GAACTGTCTC	7384
TCATGAACCT CAAGCTGCAT CTAGAGGCTT CCTTCATTTT CTCCGTCACC TCAGAGACAT	7444
ACACCTATGT CATTTCAATTT CCTATTTTGT GAAGAGGACT CCTTAAATTT GGGGGACTTA	7504
CATGATTCAT TTAAACATCT GAGAAAAGCT TTGAACCCTG GGACGTGGCT AGTCATAACC	7564
TTACCAGATT TTACACATG TATCTATGCA TTTTCTGGAC CCGTTCAACT TTTCCTTTGA	7624
ATCCTCTCTC TGTGTTACCC AGTAACTCAT CTGTCACCAA GCCTTGGGGA TTCTTCCATC	7684
TGATTGTGAT GTGAGTTGCA CAGCTATGAA GGCTGTACAC TGCACGAATG GAAGAGGCAC	7744
CTGTCCCAGA AAAAGCATCA TGGCTATCTG TGGGTAGTAT GATGGGTGTT TTTAGCAGGT	7804
AGGAGGCAAA TATCTTGAAA GGGGTGTGA AGAGGTGTTT TTTCTAATTG GCATGAAGGT	7864
GTCATACAGA TTTGCAAAGT TTAATGGTGC CTTCAATTTG GATGCTACTC TAGTATTCCA	7924
GACCTGAAGA ATCACAATAA TTTTCTACCT GGTCTCTCCT TGTCTGATA ATGAAAATTA	7984
TGATAAGGAT GATAAAAGCA CTTACTTCGT GTCCGACTCT TCTGAGCACC TACTTACATG	8044
CATTACTGCA TGCACTTCTT ACAATAATTC TATGAGATAG GTACTATTAT CCCCATTCTT	8104
TTTTTAAATG AAGAAAGTGA AGTAGGCCGG GCACGGTGCG TCACGCCTGT AATCCCAGCA	8164
CTTTGGGAGG CCAAAGCGGG TGGATCACGA GGTCAGGAGA TCGAGACCAT CCTGGCTAAC	8224
ATGGTGAAAC CCCATCTCTA ATAAAAATAC AAAAAATTAG CTGGGCGTGG TGGCAGACGC	8284
CTGTAGTCCC AGCTACTCGG AAGGCTGAGG CAGGAGAATG GCATGAACCC AGGAGGCAGA	8344
GCTTGCAGTG AGCCGAGTTT GCGCCACTGC ACTCCAGCCT AGGTGACAGA GTGAGACTCC	8404

ATCTCAAAAA	AATAAAAAATA	AAAATAAAAA	AATGAAAAAA	AAAAGAAAGT	GAAGTATAGA	8464
GTATCTCATA	GTTTGTCACT	GATAGAAACA	GGTTTCAAAC	TCAGTCAATC	TGACCGTTTG	8524
ATACATCTCA	GACACCACTA	CATTCACTAG	TTTAGATGCC	TAGAATAAAT	AGAGAAGGAA	8584
GGAGATGGCT	CTTCTCTTGT	CTCATTGTGT	TTCTTCTGAG	TGAGCTTGAA	TCACATGAAG	8644
GGGAACAGCA	GAAAACAACC	AACTGATCCT	CAGCTGTCAT	GTTTCCTTTA	AAAGTCCCTG	8704
AAGGAAGGTC	CTGGAATGTG	ACTCCCTTGC	TCTCTGTGTT	CTCTCTTTGG	CATTTCATTTT	8764
TTTGGACCCT	ACGCAAGGAC	TGTAATTGGT	GGGGACAGCT	AGTGGCCCTG	CTGGGCTTCA	8824
CACACGGTGT	CCTCCCTAGG	CCAGTGCCTC	TGGAGTCAGA	ACTCTGGTGG	TATTTCCCTC	8884
AATGAAGTGG	AGTAAGCTCT	CTCATTTTGA	GATGGTATAA	TGGAAGCCAC	CAAGTGGCTT	8944
AGAGGATGCC	CAGGTCCTTC	CATGGAGCCA	CTGGGGTTCC	GGTGACACAT	AAAAAAAAAA	9004
TCTAACCAGG	ACATTCAGGA	ATTGCTAGAT	TCTGGGAAAT	CAGTTCACCA	TGTTCAAAG	9064
AGTCTTTTTT	TTTTTTTTGA	GACTCTATTG	CCCAGGCTGG	AGTGCAATGG	CATGATCTCG	9124
GCTCACTGTA	ACCTCTGCCT	CCCAGGTTCA	AGCGATTCTC	CTGTCTCAGC	CTCCCAAGTA	9184
GCTGGGATTA	CAGGCGTGCA	CCACCATGCC	CGGCTAATTT	TTGTATTTTT	AGTAGAGACA	9244
GGGTTTCACC	ATGTTGGCCA	GGCTGGTCTC	GAAGTCTCCT	GACCTCGTGA	TCCGCCTGCC	9304
TCGGCCTCCC	AAAGTGCTGA	GATTACAGGT	GTGAGCCACC	CTGCCCAGCC	GTCAAAGAG	9364
TCTTAATATA	TATATCCAGA	TGGCATGTGT	TTACTTTATG	TTACTACATG	CACCTGGCTG	9424
CATAAATGTG	GTACAAGCAT	TCTGTCTTGA	AGGGCAGGTG	CTTCAGGATA	CCATATACAG	9484
CTCAGAAGTT	TCTTCTTTAG	GCATTAAATT	TTAGCAAAGA	TATCTCATCT	CTTCTTTTAA	9544
ACCATTTTCT	TTTTTTGTGG	TTAGAAAAGT	TATGTAGAAA	AAAGTAAATG	TGATTTACGC	9604
TCATTGTAGA	AAAGCTATAA	AATGAATACA	ATTAAAGCTG	TTATTTAATT	AGCCAGTGAA	9664
AAACTATTAA	CAACTTGTCT	ATTACCTGTT	AGTATTATTG	TTGCATTAAA	AATGCATATA	9724
CTTTAATAAA	TGTATATTGT	ATTGTATACT	GCATGATTTT	ATTGAAGTTC	TTGTTTCTCT	9784
TGTGTATATA	CTTAATCGCT	TTGTCAATTT	GGAGACATTT	ATTTTGCTTC	TAATTTCTTT	9844
ACATTTTGTC	TTACGGAATA	TTTTCAATTC	ACTGTGGTAG	CCGAATTAAT	CGTGTTCCTT	9904
CACCTAGGG	ACATTGTCGT	CTAAGTTGTA	AGACATTGGT	TATTTTACCA	GCAAACCATT	9964
CTGAAAGCAT	ATGACAAATT	ATTTCTCTCT	TAATATCTTA	CTATACTGAA	AGCAGACTGC	10024
TATAAGGCTT	CACTTACTCT	TCTACCTCAT	AAGGAATATG	TTACAATTAA	TTTATTAGGT	10084
AAGCATTTGT	TTTATATTGG	TTTTATTTCA	CCTGGGCTGA	GATTTCAAGA	AACACCCCAG	10144
TCTTCACAGT	AACACATTTT	ACTAACACAT	TTACTAAACA	TCAGCAACTG	TGGCCTGTTA	10204
ATTTTTTTAA	TAGAAATTTT	AAGTCCTCAT	TTTCTTTTCG	TGTTTTTTAA	GCTTAATTTT	10264

```

TCTGGCTTTA TTCATAAATT CTTAAGGTCA ACTACATTTG AAAAATCAAA GACCTGCATT 10324
TTAAATTCTT ATTCACCTCT GGCAAAACCA TTCACAAACC ATGGTAGTAA AGAGAAGGGT 10384
GACACCTGGT GGCCATAGGT AAATGTACCA CGGTGGTCCG GTGACCAGAG ATGCAGCGCT 10444
GAGGGTTTTT CTGAAGGTAA AGGAATAAAG AATGGGTGGA GGGGCGTGCA CTGGAATCA 10504
CTTGTAGAGA AAAGCCCCTG AAAATTTGAG AAAACAAACA AGAACTACT TACCAGCTAT 10564
TTGAATTGCT GGAATCACAG GCCATTGCTG AGCTGCCTGA ACTGGGAACA CAACAGAAGG 10624
AAAACAAACC ACTCTGATAA TCATTGAGTC AAGTACAGCA GGTGATTGAG GACTGCTGAG 10684
AGGTACAGGC CAAAATTCTT ATGTTGTATT ATAATAATGT CATCTTATAA TACTGTCAGT 10744
ATTTTATAAA ACATTCTTCA CAAACTCACA CACATTTAAA AACAAAACAC TGTCTCTAAA 10804
ATCCCCAAAT TTTTCATAAA C 10825

```

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 348 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```

Met Gly Pro Arg Ala Arg Pro Ala Leu Leu Leu Leu Met Leu Leu Gln
 1             5             10             15
Thr Ala Val Leu Gln Gly Arg Leu Leu Arg Ser His Ser Leu His Tyr
      20             25             30
Leu Phe Met Gly Ala Ser Glu Gln Asp Leu Gly Leu Ser Leu Phe Glu
      35             40             45
Ala Leu Gly Tyr Val Asp Asp Gln Leu Phe Val Phe Tyr Asp Asp Glu
      50             55             60
Ser Arg Arg Val Glu Pro Arg Thr Pro Trp Val Ser Ser Arg Ile Ser
      65             70             75             80
Ser Gln Met Trp Leu Gln Leu Ser Gln Ser Leu Lys Gly Trp Asp His
Met Phe Thr Val Asp Phe Trp Thr Ile Met Glu Asn His Asn His Ser
      100            105            110
Lys Glu Ser His Thr Leu Gln Val Ile Leu Gly Cys Glu Met Gln Glu
      115            120            125
Asp Asn Ser Thr Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly Gln Asp
      130            135            140
His Leu Glu Phe Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala Glu Pro
      145            150            155            160

```

```

Arg Ala Trp Pro Thr Lys Leu Glu Trp Glu Arg His Lys Ile Arg Ala
      165                                170                                175

Arg Gln Asn Arg Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln Leu Gln
      180                                185                                190

Gln Leu Leu Glu Leu Gly Arg Gly Val Leu Asp Gln Gln Val Pro Pro
      195                                200                                205

Leu Val Lys Val Thr His His Val Thr Ser Ser Val Thr Thr Leu Arg
      210                                215                                220

Cys Arg Ala Leu Asn Tyr Tyr Pro Gln Asn Ile Thr Met Lys Trp Leu
      225                                230                                235                                240

Lys Asp Lys Gln Pro Met Asp Ala Lys Glu Phe Glu Pro Lys Asp Val
      245                                250                                255

Leu Pro Asn Gly Asp Gly Thr Tyr Gln Gly Trp Ile Thr Leu Ala Val
      260                                265                                270

Pro Pro Gly Glu Glu Gln Arg Tyr Thr Tyr Gln Val Glu His Pro Gly
      275                                280                                285

Leu Asp Gln Pro Leu Ile Val Ile Trp Glu Pro Ser Pro Ser Gly Thr
      290                                295                                300

Leu Val Ile Gly Val Ile Ser Gly Ile Ala Val Phe Val Val Ile Leu
      305                                310                                315                                320

Phe Ile Gly Ile Leu Phe Ile Ile Leu Arg Lys Arg Gln Gly Ser Arg
      325                                330                                335

Gly Ala Met Gly His Tyr Val Leu Ala Glu Arg Glu
      340                                345

```

## (2) INFORMATION FOR SEQ ID NO:9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1440 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 222..1268

## (ix) FEATURE:

- (A) NAME/KEY: allele
- (B) LOCATION: replace(408, "c")
- (D) OTHER INFORMATION: /phenotype= "normal or wild-type  
(unaffected)"  
/label= 24d2

## (ix) FEATURE:

(A) NAME/KEY: allele  
 (B) LOCATION: replace(414, "a")  
 (D) OTHER INFORMATION: /phenotype= "normal or wild-type  
 (unaffected)"  
 /label= 24d7

## (ix) FEATURE:

(A) NAME/KEY: allele  
 (B) LOCATION: replace(1066, "g")  
 (D) OTHER INFORMATION: /phenotype= "normal or wild-type  
 (unaffected)"  
 /label= 24d1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```

GGGGACACTG GATCACCTAG TGTTTCACAA GCAGGTACCT TCTGCTGTAG GAGAGAGAGA      60
ACTAAAGTTC TGAAAGACCT GTTGCTTTTC ACCAGGAAGT TTTACTGGGC ATCTCCTGAG      120
CCTAGGCAAT AGCTGTAGGG TGACTTCTGG AGCCATCCCC GTTTCCTCCG CCCCCAAAAG      180
AAGCGGAGAT TTAACGGGGA CGTGCGGCCA GAGCTGGGGA A ATG GGC CCG CGA      233
                               Met Gly Pro Arg
                               1

GCC AGG CCG GCG CTT CTC CTC CTG ATG CTT TTG CAG ACC GCG GTC CTG      281
Ala Arg Pro Ala Leu Leu Leu Leu Met Leu Leu Gln Thr Ala Val Leu
  5                      10                      15                      20

CAG GGG CGC TTG CTG CGT TCA CAC TCT CTG CAC TAC CTC TTC ATG GGT      329
Gln Gly Arg Leu Leu Arg Ser His Ser Leu His Tyr Leu Phe Met Gly
                      25                      30                      35

GCC TCA GAG CAG GAC CTT GGT CTT TCC TTG TTT GAA GCT TTG GGC TAC      377
Ala Ser Glu Gln Asp Leu Gly Leu Ser Leu Phe Glu Ala Leu Gly Tyr
                      40                      45                      50

GTG GAT GAC CAG CTG TTC GTG TTC TAT GAT CAT GAG AGT CGC CGT GTG      425
Val Asp Asp Gln Leu Phe Val Phe Tyr Asp His Glu Ser Arg Arg Val
                      55                      60                      65

GAG CCC CGA ACT CCA TGG GTT TCC AGT AGA ATT TCA AGC CAG ATG TGG      473
Glu Pro Arg Thr Pro Trp Val Ser Ser Arg Ile Ser Ser Gln Met Trp
                      70                      75                      80

CTG CAG CTG AGT CAG AGT CTG AAA GGG TGG GAT CAC ATG TTC ACT GTT      521
Leu Gln Leu Ser Gln Ser Leu Lys Gly Trp Asp His Met Phe Thr Val
                      85                      90                      95                      100

GAC TTC TGG ACT ATT ATG GAA AAT CAC AAC CAC AGC AAG GAG TCC CAC      569
Asp Phe Trp Thr Ile Met Glu Asn His Asn His Ser Lys Glu Ser His
                      105                      110                      115

ACC CTG CAG GTC ATC CTG GGC TGT GAA ATG CAA GAA GAC AAC AGT ACC      617
Thr Leu Gln Val Ile Leu Gly Cys Glu Met Gln Glu Asp Asn Ser Thr
                      120                      125                      130

GAG GGC TAC TGG AAG TAC GGG TAT GAT GGG CAG GAC CAC CTT GAA TTC      663
Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly Gln Asp His Leu Glu Phe

```

135	140	145	
TGC CCT GAC ACA CTG GAT TGG AGA GCA GCA GAA CCC AGG GCC TGG CCC			713
Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala Glu Pro Arg Ala Trp Pro			
150	155	160	
ACC AAG CTG GAG TGG GAA AGG CAC AAG ATT CGG GCC AGG CAG AAC AGG			761
Thr Lys Leu Glu Trp Glu Arg His Lys Ile Arg Ala Arg Gln Asn Arg			
165	170	175	180
GCC TAC CTG GAG AGG GAC TGC CCT GCA CAG CTG CAG CAG TTG CTG GAG			809
Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln Leu Gln Gln Leu Leu Glu			
	185	190	195
CTG GGG AGA GGT GTT TTG GAC CAA CAA GTG CCT CCT TTG GTG AAG GTG			857
Leu Gly Arg Gly Val Leu Asp Gln Gln Val Pro Pro Leu Val Lys Val			
	200	205	210
ACA CAT CAT GTG ACC TCT TCA GTG ACC ACT CTA CGG TGT CGG GCC TTG			905
Thr His His Val Thr Ser Ser Val Thr Thr Leu Arg Cys Arg Ala Leu			
	215	220	225
AAC TAC TAC CCC CAG AAC ATC ACC ATG AAG TGG CTG AAG GAT AAG CAG			953
Asn Tyr Tyr Pro Gln Asn Ile Thr Met Lys Trp Leu Lys Asp Lys Gln			
	230	235	240
CCA ATG GAT GCC AAG GAG TTC GAA CCT AAA GAC GTA TTG CCC AAT GGG			1001
Pro Met Asp Ala Lys Glu Phe Glu Pro Lys Asp Val Leu Pro Asn Gly			
	245	250	255
GAT GGG ACC TAC CAG GGC TGG ATA ACC TTG GCT GTA CCC CCT GGG GAA			1049
Asp Gly Thr Tyr Gln Gly Trp Ile Thr Leu Ala Val Pro Pro Gly Glu			
	265	270	275
GAG CAG AGA TAT ACG TGC CAG GTG GAG CAC CCA GGC CTG GAT CAG CCC			1097
Glu Gln Arg Tyr Thr Cys Gln Val Glu His Pro Gly Leu Asp Gln Pro			
	280	285	290
CTC ATT GTG ATC TGG GAG CCC TCA CCG TCT GGC ACC CTA GTC ATT GGA			1145
Leu Ile Val Ile Trp Glu Pro Ser Pro Ser Gly Thr Leu Val Ile Gly			
	295	300	305
GTC ATC AGT GGA ATT GCT GTT TTT GTC GTC ATC TTG TTC ATT GGA ATT			1193
Val Ile Ser Gly Ile Ala Val Phe Val Val Ile Leu Phe Ile Gly Ile			
	310	315	320
TTG TTC ATA ATA TTA AGG AAG AGG CAG GGT TCA AGA GGA GCC ATG GGG			1241
Leu Phe Ile Ile Leu Arg Lys Arg Gln Gly Ser Arg Gly Ala Met Gly			
	325	330	335
CAC TAC GTC TTA GCT GAA CGT GAG TGACACGCAG CCTGCAGACT CACTGTGGGA			1295
His Tyr Val Leu Ala Glu Arg Glu			
	345		
AGGAGACAAA ACTAGAGACT CAAAGAGGGA GTGCATTTAT GAGCTCTTCA TGTTCAGGA			1355
GAGAGTTGAA CCTAAACATA GAAATGCGCT GACGAACCTC TTGATTTTAG CCTTCTCTGT			1415
TCATTCCTC AAAAAGATTT CCCC			1440

## (2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1440 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

- (ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION: 222..1268

- (ix) FEATURE:  
 (A) NAME/KEY: allele  
 (B) LOCATION: replace(1066, "a")  
 (D) OTHER INFORMATION: /phenotype= "Hereditary Hemochromatosis (HH)"  
 /label= 24d1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```

GGGGACACTG GATCACCTAG TGTTTCACAA GCAGGTACCT TCTGCTGTAG GAGAGAGAGA      60
ACTAAAGTTC TGAAAGACCT GTTGCTTTTC ACCAGGAAGT TTTACTGGGC ATCTCCTGAG      120
CCTAGGCAAT AGCTGTAGGG TGACTTCTGG AGCCATCCCC GTTTCCTCCG CCCCCAAAAG      180
AAGCGGAGAT TTAACGGGGA CGTGCGGCCA GAGCTGGGGA A ATG GGC CCG CGA      233
                                     Met Gly Pro Arg
                                     1

GCC AGG CCG GCG CTT CTC CTC CTG ATG CTT TTG CAG ACC GCG GTC CTG      281
Ala Arg Pro Ala Leu Leu Leu Leu Met Leu Leu Gln Thr Ala Val Leu
   5                      10                      15                      20

CAG GGG CGC TTG CTG CGT TCA CAC TCT CTG CAC TAC CTC TTC ATG GGT      329
Gln Gly Arg Leu Leu Arg Ser His Ser Leu His Tyr Leu Phe Met Gly
                      25                      30                      35

GCC TCA GAG CAG GAC CTT GGT CTT TCC TTG TTT GAA GCT TTG GGC TAC      377
Ala Ser Glu Gln Asp Leu Gly Leu Ser Leu Phe Glu Ala Leu Gly Tyr
                      40                      45                      50

GTG GAT GAC CAG CTG TTC GTG TTC TAT GAT CAT GAG AGT CGC CGT GTG      425
Val Asp Asp Gln Leu Phe Val Phe Tyr Asp His Glu Ser Arg Arg Val
                      55                      60                      65

GAG CCC CGA ACT CCA TGG GTT TCC AGT AGA ATT TCA AGC CAG ATG TGG      473
Glu Pro Arg Thr Pro Trp Val Ser Ser Arg Ile Ser Ser Gln Met Trp
                      70                      75                      80

CTG CAG CTG AGT CAG AGT CTG AAA GGG TGG GAT CAC ATG TTC ACT GTT      521
Leu Gln Leu Ser Gln Ser Leu Lys Gly Trp Asp His Met Phe Thr Val
                      85                      90                      95                      100

GAC TTC TGG ACT ATT ATG GAA AAT CAC AAC CAC AGC AAG GAG TCC CAC      569

```

Asp Phe Trp Thr Ile Met Glu Asn His Asn His Ser Lys Glu Ser His	105	110	115	
ACC CTG CAG GTC ATC CTG GGC TGT GAA ATG CAA GAA GAC AAC AGT ACC				617
Thr Leu Gln Val Ile Leu Gly Cys Glu Met Gln Glu Asp Asn Ser Thr	120	125	130	
GAG GGC TAC TGG AAG TAC GGG TAT GAT GGG CAG GAC CAC CTT GAA TTC				665
Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly Gln Asp His Leu Glu Phe	135	140	145	
TGC CCT GAC ACA CTG GAT TGG AGA GCA GCA GAA CCC AGG GCC TGG CCC				713
Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala Glu Pro Arg Ala Trp Pro	150	155	160	
ACC AAG CTG GAG TGG GAA AGG CAC AAG ATT CGG GCC AGG CAG AAC AGG				761
Thr Lys Leu Glu Trp Glu Arg His Lys Ile Arg Ala Arg Gln Asn Arg	165	170	175	180
GCC TAC CTG GAG AGG GAC TGC CCT GCA CAG CTG CAG CAG TTG CTG GAG				809
Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln Leu Gln Gln Leu Leu Glu	185	190	195	
CTG GGG AGA GGT GTT TTG GAC CAA CAA GTG CCT CCT TTG GTG AAG GTG				857
Leu Gly Arg Gly Val Leu Asp Gln Gln Val Pro Pro Leu Val Lys Val	200	205	210	
ACA CAT CAT GTG ACC TCT TCA GTG ACC ACT CTA CGG TGT CGG GCC TTG				905
Thr His His Val Thr Ser Ser Val Thr Thr Leu Arg Cys Arg Ala Leu	215	220	225	
AAC TAC TAC CCC CAG AAC ATC ACC ATG AAG TGG CTG AAG GAT AAG CAG				953
Asn Tyr Tyr Pro Gln Asn Ile Thr Met Lys Trp Leu Lys Asp Lys Gln	230	235	240	
CCA ATG GAT GCC AAG GAG TTC GAA CCT AAA GAC GTA TTG CCC AAT GGG				1001
Pro Met Asp Ala Lys Glu Phe Glu Pro Lys Asp Val Leu Pro Asn Gly	245	250	255	260
GAT GGG ACC TAC CAG GGC TGG ATA ACC TTG GCT GTA CCC CCT GGG GAA				1049
Asp Gly Thr Tyr Gln Gly Trp Ile Thr Leu Ala Val Pro Pro Gly Glu	265	270	275	
GAG CAG AGA TAT ACG TAC CAG GTG GAG CAC CCA GGC CTG GAT CAG CCC				1097
Glu Gln Arg Tyr Thr Tyr Gln Val Glu His Pro Gly Leu Asp Gln Pro	280	285	290	
CTC ATT GTG ATC TGG GAG CCC TCA CCG TCT GGC ACC CTA GTC ATT GGA				1145
Leu Ile Val Ile Trp Glu Pro Ser Pro Ser Gly Thr Leu Val Ile Gly	295	300	305	
GTC ATC AGT GGA ATT GCT GTT TTT GTC GTC ATC TTG TTC ATT GGA ATT				1193
Val Ile Ser Gly Ile Ala Val Phe Val Val Ile Leu Phe Ile Gly Ile	310	315	320	
TTG TTC ATA ATA TTA AGG AAG AGG CAG GGT TCA AGA GGA GCC ATG GGG				1241
Leu Phe Ile Ile Leu Arg Lys Arg Gln Gly Ser Arg Gly Ala Met Gly	325	330	335	340
CAC TAC GTC TTA GCT GAA CGT GAG TGACACGCAG CCTGCAGACT CACTGTGGGA				1295
His Tyr Val Leu Ala Glu Arg Glu				

345

AGGAGACAAA	ACTAGAGACT	CAAAGAGGGA	GTGCATTTAT	GAGCTCTTCA	TGTTTCAGGA	1355
GAGAGTTGAA	CCTAAACATA	GAAATTGCCT	GACGAACTCC	TTGATTTTAG	CCTTCTCTGT	1415
TCATTTCCCTC	AAAAAGATTT	CCCCA				1440

(2) INFORMATION FOR SEQ ID NO:11:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1440 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: CDS  
(B) LOCATION: 222..1268

(ix) FEATURE:

- ```
(A) NAME/KEY: allele  
(B) LOCATION: replace(408, "g")  
(D) OTHER INFORMATION: /phenotype= "Hereditary Hemochromatosis  
                        (HH)"  
                        /label= 24d2
```

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

|                                                                 |            |            |            |                   |            |    |       |
|-----------------------------------------------------------------|------------|------------|------------|-------------------|------------|----|-------|
| GGGGACACTG                                                      | GATCACCTAG | TGTTTCACAA | GCAGGTACCT | TCTGCTGTAG        | GAGAGAGAGA |    | 60    |
| ACTAAAGTTC                                                      | TGAAAGACCT | GTTGCTTTTC | ACCAGGAAGT | TTTACTGGGC        | ATCTCCTGAG |    | 120   |
| CCTAGGCAAT                                                      | AGCTGTAGGG | TGACTTCTGG | AGCCATCCCC | GTTTCCCCGC        | CCCCAAAAG  |    | 180   |
| AAGCGGAGAT                                                      | TTAACGGGGA | CGTGCGGCCA | GAGCTGGGGA | A ATG GGC CCG CGA |            |    | 233   |
|                                                                 |            |            |            | Met Gly Pro Arg   |            |    |       |
|                                                                 |            |            |            | 1                 |            |    |       |
| GCC AGG CCG GCG CTT CTC CTC CTG ATG CTT TTG CAG ACC GCG GTC CTG |            |            |            |                   |            |    | 281   |
| Ala Arg Pro Ala Leu Leu Leu Leu Met Leu Leu Gln Thr Ala Val Leu |            |            |            |                   |            |    |       |
| 5                                                               |            |            |            | 10                |            |    | 15 20 |
| CAG GGG CGC TTG CTG CGT TCA CAC TCT CTG CAC TAC CTC TTC ATG GGT |            |            |            |                   |            |    | 329   |
| Gln Gly Arg Leu Leu Arg Ser His Ser Leu His Tyr Leu Phe Met Gly |            |            |            |                   |            |    |       |
|                                                                 |            |            | 25         |                   |            | 30 | 35    |
| GCC TCA GAG CAG GAC CTT GGT CTT TCC TTG TTT GAA GCT TTG GGC TAC |            |            |            |                   |            |    | 377   |
| Ala Ser Glu Gln Asp Leu Gly Leu Ser Leu Phe Glu Ala Leu Gly Tyr |            |            |            |                   |            |    |       |
|                                                                 |            | 40         |            |                   |            | 45 | 50    |
| GTG GAT GAC CAG CTG TTC GTG TTC TAT GAT GAT GAG AGT CGC CGT GTG |            |            |            |                   |            |    | 425   |
| Val Asp Asp Gln Leu Phe Val Phe Tyr Asp Asp Glu Ser Arg Arg Val |            |            |            |                   |            |    |       |
|                                                                 | 55         |            |            |                   | 60         |    | 65    |

|                                                                                                                                                       |      |
|-------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| GAG CCC CGA ACT CCA TGG GTT TCC AGT AGA ATT TCA AGC CAG ATG TGG<br>Glu Pro Arg Thr Pro Trp Val Ser Ser Arg Ile Ser Ser Gln Met Trp<br>70 75 80        | 473  |
| CTG CAG CTG AGT CAG AGT CTG AAA GGG TGG GAT CAC ATG TTC ACT GTT<br>Leu Gln Leu Ser Gln Ser Leu Lys Gly Trp Asp His Met Phe Thr Val<br>85 90 95 100    | 521  |
| GAC TTC TGG ACT ATT ATG GAA AAT CAC AAC CAC AGC AAG GAG TCC CAC<br>Asp Phe Trp Thr Ile Met Glu Asn His Asn His Ser Lys Glu Ser His<br>105 110 115     | 569  |
| ACC CTG CAG GTC ATC CTG GGC TGT GAA ATG CAA GAA GAC AAC AGT ACC<br>Thr Leu Gln Val Ile Leu Gly Cys Glu Met Gln Glu Asp Asn Ser Thr<br>120 125 130     | 617  |
| GAG GGC TAC TGG AAG TAC GGG TAT GAT GGG CAG GAC CAC CTT GAA TTC<br>Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly Gln Asp His Leu Glu Phe<br>135 140 145     | 665  |
| TGC CCT GAC ACA CTG GAT TGG AGA GCA GCA GAA CCC AGG GCC TGG CCC<br>Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala Glu Pro Arg Ala Trp Pro<br>150 155 160     | 713  |
| ACC AAG CTG GAG TGG GAA AGG CAC AAG ATT CGG GCC AGG CAG AAC AGG<br>Thr Lys Leu Glu Trp Glu Arg His Lys Ile Arg Ala Arg Gln Asn Arg<br>165 170 175 180 | 761  |
| GCC TAC CTG GAG AGG GAC TGC CCT GCA CAG CTG CAG CAG TTG CTG GAG<br>Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln Leu Gln Gln Leu Leu Glu<br>185 190 195     | 809  |
| CTG GGG AGA GGT GTT TTG GAC CAA CAA GTG CCT CCT TTG GTG AAG GTG<br>Leu Gly Arg Gly Val Leu Asp Gln Gln Val Pro Pro Leu Val Lys Val<br>200 205 210     | 857  |
| ACA CAT CAT GTG ACC TCT TCA GTG ACC ACT CTA CGG TGT CGG GCC TTG<br>Thr His His Val Thr Ser Ser Val Thr Thr Leu Arg Cys Arg Ala Leu<br>215 220 225     | 905  |
| AAC TAC TAC CCC CAG AAC ATC ACC ATG AAG TGG CTG AAG GAT AAG CAG<br>Asn Tyr Tyr Pro Gln Asn Ile Thr Met Lys Trp Leu Lys Asp Lys Gln<br>230 235 240     | 953  |
| CCA ATG GAT GCC AAG GAG TTC GAA CCT AAA GAC GTA TTG CCC AAT GGG<br>Pro Met Asp Ala Lys Glu Phe Glu Pro Lys Asp Val Leu Pro Asn Gly<br>245 250 255 260 | 1001 |
| GAT GGG ACC TAC CAG GGC TGG ATA ACC TTG GCT GTA CCC CCT GGG GAA<br>Asp Gly Thr Tyr Gln Gly Trp Ile Thr Leu Ala Val Pro Pro Gly Glu<br>265 270 275     | 1049 |
| GAG CAG AGA TAT ACG TGC CAG GTG GAG CAC CCA GGC CTG GAT CAG CCC<br>Glu Gln Arg Tyr Thr Cys Gln Val Glu His Pro Gly Leu Asp Gln Pro<br>280 285 290     | 1097 |
| CTC ATT GTG ATC TGG GAG CCC TCA CCG TCT GGC ACC CTA GTC ATT GGA<br>Leu Ile Val Ile Trp Glu Pro Ser Pro Ser Gly Thr Leu Val Ile Gly<br>295 300 305     | 1145 |
| GTC ATC AGT GGA ATT GCT GTT TTT GTC GTC ATC TTG TTC ATT GGA ATT                                                                                       | 1193 |

```

Val Ile Ser Gly Ile Ala Val Phe Val Val Ile Leu Phe Ile Gly Ile
  310                      315                      320

TTG TTC ATA ATA TTA AGG AAG AGG CAG GGT TCA AGA GGA GCC ATG GGG      1241
Leu Phe Ile Ile Leu Arg Lys Arg Gln Gly Ser Arg Gly Ala Met Gly
  325                      330                      335                      340

CAC TAC GTC TTA GCT GAA CGT GAG TGACACGCAG CCTGCAGACT CACTGTGGGA      1295
His Tyr Val Leu Ala Glu Arg Glu
                      345

AGGAGACAAA ACTAGAGACT CAAAGAGGGA GTGCATTTAT GAGCTCTTCA TGTTTCAGGA      1355

GAGAGTTGAA CCTAAACATA GAAATTGCCT GACGAACTCC TTGATTTTAG CCTTCTCTGT      1415

TCATTTCTCTC AAAAAGATTT CCCCCA  1440

```

## (2) INFORMATION FOR SEQ ID NO:12:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1440 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: cDNA

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 222..1268

## (ix) FEATURE:

- (A) NAME/KEY: allele
- (B) LOCATION: replace(408, "g")
- (D) OTHER INFORMATION: /phenotype= "Hereditary Hemochromatosis (HH)"  
/label= 24d2

## (ix) FEATURE:

- (A) NAME/KEY: allele
- (B) LOCATION: replace(1066, "a")
- (D) OTHER INFORMATION: /phenotype= "Hereditary Hemochromatosis (HH)"  
/label= 24d1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```

GGGGACACTG GATCACCTAG TGTTTCACAA GCAGGTACCT TCTGCTGTAG GAGAGAGAGA      60
ACTAAAGTTC TGAAAGACCT GTTGCTTTTC ACCAGGAAGT TTTACTGGGC ATCTCCTGAG      120
CCTAGGCAAT AGCTGTAGGG TGACTTCTGG AGCCATCCCC GTTCCCCCGC CCCCCAAAAG      180
AAGCGGAGAT TTAACGGGGA CGTGCGGCCA GAGCTGGGGA A ATG GGC CCG CGA      233
                      Met Gly Pro Arg
                      1

```

|                                                                                                                                                       |      |
|-------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| GCC AGG CCG GCG CTT CTC CTC CTG ATG CTT TTG CAG ACC GCG GTC CTG<br>Ala Arg Pro Ala Leu Leu Leu Leu Met Leu Leu Gln Thr Ala Val Leu<br>5 10 15 20      | 281  |
| CAG GGG CGC TTG CTG CGT TCA CAC TCT CTG CAC TAC CTC TTC ATG GGT<br>Gln Gly Arg Leu Leu Arg Ser His Ser Leu His Tyr Leu Phe Met Gly<br>25 30 35        | 329  |
| GCC TCA GAG CAG GAC CTT GGT CTT TCC TTG TTT GAA GCT TTG GGC TAC<br>Ala Ser Glu Gln Asp Leu Gly Leu Ser Leu Phe Glu Ala Leu Gly Tyr<br>40 45 50        | 377  |
| GTG GAT GAC CAG CTG TTC GTG TTC TAT GAT GAT GAG AGT CGC CGT GTG<br>Val Asp Asp Gln Leu Phe Val Phe Tyr Asp Asp Glu Ser Arg Arg Val<br>55 60 65        | 425  |
| GAG CCC CGA ACT CCA TGG GTT TCC AGT AGA ATT TCA AGC CAG ATG TGG<br>Glu Pro Arg Thr Pro Trp Val Ser Ser Arg Ile Ser Ser Gln Met Trp<br>70 75 80        | 473  |
| CTG CAG CTG AGT CAG AGT CTG AAA GGG TGG GAT CAC ATG TTC ACT GTT<br>Leu Gln Leu Ser Gln Ser Leu Lys Gly Trp Asp His Met Phe Thr Val<br>85 90 95 100    | 521  |
| GAC TTC TGG ACT ATT ATG GAA AAT CAC AAC CAC AGC AAG GAG TCC CAC<br>Asp Phe Trp Thr Ile Met Glu Asn His Asn His Ser Lys Glu Ser His<br>105 110 115     | 569  |
| ACC CTG CAG GTC ATC CTG GGC TGT GAA ATG CAA GAA GAC AAC AGT ACC<br>Thr Leu Gln Val Ile Leu Gly Cys Glu Met Gln Glu Asp Asn Ser Thr<br>120 125 130     | 617  |
| GAG GGC TAC TGG AAG TAC GGG TAT GAT GGG CAG GAC CAC CTT GAA TTC<br>Glu Gly Tyr Trp Lys Tyr Gly Tyr Asp Gly Gln Asp His Leu Glu Phe<br>135 140 145     | 665  |
| TGC CCT GAC ACA CTG GAT TGG AGA GCA GCA GAA CCC AGG GCC TGG CCC<br>Cys Pro Asp Thr Leu Asp Trp Arg Ala Ala Glu Pro Arg Ala Trp Pro<br>150 155 160     | 713  |
| ACC AAG CTG GAG TGG GAA AGG CAC AAG ATT CGG GCC AGG CAG AAC AGG<br>Thr Lys Leu Glu Trp Glu Arg His Lys Ile Arg Ala Arg Gln Asn Arg<br>165 170 175 180 | 761  |
| GCC TAC CTG GAG AGG GAC TGC CCT GCA CAG CTG CAG CAG TTG CTG GAG<br>Ala Tyr Leu Glu Arg Asp Cys Pro Ala Gln Leu Gln Gln Leu Leu Glu<br>185 190 195     | 809  |
| CTG GGG AGA GGT GTT TTG GAC CAA CAA GTG CCT CCT TTG GTG AAG GTG<br>Leu Gly Arg Gly Val Leu Asp Gln Gln Val Pro Pro Leu Val Lys Val<br>200 205 210     | 857  |
| ACA CAT CAT GTG ACC TCT TCA GTG ACC ACT CTA CGG TGT CGG GCC TTG<br>Thr His His Val Thr Ser Ser Val Thr Thr Leu Arg Cys Arg Ala Leu<br>215 220 225     | 905  |
| AAC TAC TAC CCC CAG AAC ATC ACC ATG AAG TGG CTG AAG GAT AAG CAG<br>Asn Tyr Tyr Pro Gln Asn Ile Thr Met Lys Trp Leu Lys Asp Lys Gln<br>230 235 240     | 953  |
| CCA ATG GAT GCC AAG GAG TTC GAA CCT AAA GAC GTA TTG CCC AAT GGG                                                                                       | 1001 |

|                                                                   |      |
|-------------------------------------------------------------------|------|
| Pro Met Asp Ala Lys Glu Phe Glu Pro Lys Asp Val Leu Pro Asn Gly   |      |
| 245 250 255 260                                                   |      |
| GAT GGG ACC TAC CAG GGC TGG ATA ACC TTG GCT GTA CCC CCT GGG GAA   | 1049 |
| Asp Gly Thr Tyr Gln Gly Trp Ile Thr Leu Ala Val Pro Pro Gly Glu   |      |
| 265 270 275                                                       |      |
| GAG CAG AGA TAT ACG TAC CAG GTG GAG CAC CCA GGC CTG GAT CAG CCC   | 1097 |
| Glu Gln Arg Tyr Thr Tyr Gln Val Glu His Pro Gly Leu Asp Gln Pro   |      |
| 280 285 290                                                       |      |
| CTC ATT GTG ATC TGG GAG CCC TCA CCG TCT GGC ACC CTA GTC ATT GGA   | 1145 |
| Leu Ile Val Ile Trp Glu Pro Ser Pro Ser Gly Thr Leu Val Ile Gly   |      |
| 295 300 305                                                       |      |
| GTC ATC AGT GGA ATT GCT GTT TTT GTC GTC ATC TTG TTC ATT GGA ATT   | 1193 |
| Val Ile Ser Gly Ile Ala Val Phe Val Val Ile Leu Phe Ile Gly Ile   |      |
| 310 315 320                                                       |      |
| TTG TTC ATA ATA TTA AGG AAG AGG CAG GGT TCA AGA GGA GCC ATG GGG   | 1241 |
| Leu Phe Ile Ile Leu Arg Lys Arg Gln Gly Ser Arg Gly Ala Met Gly   |      |
| 325 330 335 340                                                   |      |
| CAC TAC GTC TTA GCT GAA CGT GAG TGACACGCAG CCTGCAGACT CACTGTGGGA  | 1295 |
| His Tyr Val Leu Ala Glu Arg Glu                                   |      |
| 345                                                               |      |
| AGGAGACAAA ACTAGAGACT CAAAGAGGGA GTGCATTTAT GAGCTCTTCA TGTTTCAGGA | 1355 |
| GAGAGTTGAA CCTAAACATA GAAATTGCCT GACGAACTCC TTGATTTTAG CCTTCTCTGT | 1415 |
| TCATTTCCCTC AAAAAGATTT CCCCCA                                     | 1440 |

## (2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TGCCAAGGGT AACAGATCC

20

## (2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CTCAGGCACT CCTCTCAACC

20

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-biotinylated guanine  
(bio-G)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

NGAAGAGCAG AGATATACGT G

21

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-biotinylated guanine  
(bio-G)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

NGAAGAGCAG AGATATACGT A

21

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-phosphorylated cytosine  
(p-C) "

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 18
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 3'-digoxigenin-conjugated  
guanine (G-dig) "

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

NCAGGTGGAG CACCCAGN

18

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTGAAAGGGT GGGATCACAT

20

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CAAGGAGTTC GTCAGGCAAT

20

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 517 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: -

(B) LOCATION: 1..517

(D) OTHER INFORMATION: /note= "normal or wild-type (unaffected) genomic sequence surrounding variant for 24d1(G) allele corresponding to positions 5507-6023 of genomic sequence containing the HH gene (SEQ ID NO:1)"

(ix) FEATURE:

(A) NAME/KEY: allele

(B) LOCATION: replace(328, "g")

(D) OTHER INFORMATION: /phenotype= "normal or wild-type (unaffected)"  
/label= 24d1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| TATTTCTTC CTCCAACCTA TAGAAGGAAG TGAAAGTTCC AGTCTTCCTG GCAAGGGTAA   | 60  |
| ACAGATCCCC TCTCCTCATC CTTCCTCTTT CCTGTCAAGT GCCTCCTTTG GTGAAGGTGA  | 120 |
| CACATCATGT GACCTCTTCA GTGACCACTC TACGGTGTCTG GGCCTTGAAC TACTACCCCC | 180 |
| AGAACATCAC CATGAAGTGG CTGAAGGATA AGCAGCCAAT GGATGCCAAG GAGTTCGAAC  | 240 |
| CTAAAGACGT ATTGCCCAAT GGGGATGGGA CCTACCAGGG CTGGATAACC TTGGCTGTAC  | 300 |
| CCCCTGGGGA AGAGCAGAGA TATACGTGCC AGGTGGAGCA CCCAGGCCTG GATCAGCCCC  | 360 |
| TCATTGTGAT CTGGGGTATG TGAATGATGA GAGCCAGGAG CTGAGAAAAT CTATTGGGGG  | 420 |
| TTGAGAGGAG TGCCTGAGGA GGTAATTATG GCAGTGAGAT GAGGATCTGC TCTTTGTTAG  | 480 |
| GGGGTGGGCT GAGGGTGGCA ATCAAAGGCT TTAAGTT                           | 517 |

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 517 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: -

(B) LOCATION: 1..517

(D) OTHER INFORMATION: /note= "genomic sequence surrounding variant for 24d1(A) allele corresponding to positions 5507-6023 of genomic sequence containing the HH gene"

(SEQ ID NO:3)"

## (ix) FEATURE:

- (A) NAME/KEY: allele
- (B) LOCATION: replace(328, "a")
- (D) OTHER INFORMATION: /phenotype= "Hereditary Hemochromatosis (HH)"  
/label= 24d1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

```

TATTTCTTC CTCCAACCTA TAGAAGGAAG TGAAAGTTCC AGTCTTCCTG GCAAGGGTAA      60
ACAGATCCCC TCTCCTCATC CTTCTCTTT CCTGTCAAGT GCCTCCTTTG GTGAAGGTGA      120
CACATCATGT GACCTCTTCA GTGACCACTC TACGGTGTCTG GGCCTTGAAC TACTACCCCC      180
AGAACATCAC CATGAAGTGG CTGAAGGATA AGCAGCCAAT GGATGCCAAG GAGTTCGAAC      240
CTAAAGACGT ATTGCCCAAT GGGGATGGGA CCTACCAGGG CTGGATAACC TTGGCTGTAC      300
CCCCTGGGGA AGAGCAGAGA TATACGTACC AGGTGGAGCA CCCAGGCCTG GATCAGCCCC      360
TCATTGTGAT CTGGGGTATG TGA CTGATGA GAGCCAGGAG CTGAGAAAAT CTATTGGGGG      420
TTGAGAGGAG TGCCTGAGGA GGTAATTATG GCAGTGAGAT GAGGATCTGC TCTTTGTTAG      480
GGGGTGGGCT GAGGGTGGCA ATCAAAGGCT TTA ACTT                               517

```

## (2) INFORMATION FOR SEQ ID NO:22:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 361 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (ix) FEATURE:

- (A) NAME/KEY: Protein
- (B) LOCATION: 1..361
- (D) OTHER INFORMATION: /note= "Rabbit leukocyte antigen (RLA)"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

```

Met Gly Ser Ile Pro Pro Arg Thr Leu Leu Leu Leu Leu Ala Gly Ala .
1           5           10           15
Leu Thr Leu Lys Asp Thr Gln Ala Gly Ser His Ser Met Arg Tyr Phe
20           25           30
Tyr Thr Ser Val Ser Arg Pro Gly Leu Gly Glu Pro Arg Phe Ile Ile
35           40           45
Val Gly Tyr Val Asp Asp Thr Gln Phe Val Arg Phe Asp Ser Asp Ala
50           55           60

```

Ala Ser Pro Arg Met Glu Gln Arg Ala Pro Trp Met Gly Gln Val Glu  
 65 70 75 80  
 Pro Glu Tyr Trp Asp Gln Gln Thr Gln Ile Ala Lys Asp Thr Ala Gln  
 85 90 95  
 Thr Phe Arg Val Asn Leu Asn Thr Ala Leu Arg Tyr Tyr Asn Gln Ser  
 100 105 110  
 Ala Ala Gly Ser His Thr Phe Gln Thr Met Phe Gly Cys Glu Val Trp  
 115 120 125  
 Ala Asp Gly Arg Phe Phe His Gly Tyr Arg Gln Tyr Ala Tyr Asp Gly  
 130 135 140  
 Ala Asp Tyr Ile Ala Leu Asn Glu Asp Leu Arg Ser Trp Thr Ala Ala  
 145 150 155 160  
 Asp Thr Ala Ala Gln Asn Thr Gln Arg Lys Trp Glu Ala Ala Gly Glu  
 165 170 175  
 Ala Glu Arg His Arg Ala Tyr Leu Glu Arg Glu Cys Val Glu Trp Leu  
 180 185 190  
 Arg Arg Tyr Leu Glu Met Gly Lys Glu Thr Leu Gln Arg Ala Asp Pro  
 195 200 205  
 Pro Lys Ala His Val Thr His His Pro Ala Ser Asp Arg Glu Ala Thr  
 210 215 220  
 Leu Arg Cys Trp Ala Leu Gly Phe Tyr Pro Ala Glu Ile Ser Leu Thr  
 225 230 235 240  
 Trp Gln Arg Asp Gly Glu Asp Gln Thr Gln Asp Thr Glu Leu Val Glu  
 245 250 255  
 Thr Arg Pro Gly Gly Asp Gly Thr Phe Gln Lys Trp Ala Ala Val Val  
 260 265 270  
 Val Pro Ser Gly Glu Glu Gln Arg Tyr Thr Cys Arg Val Gln His Glu  
 275 280 285  
 Gly Leu Pro Glu Pro Leu Thr Leu Thr Trp Glu Pro Pro Ala Gln Pro  
 290 295 300  
 Thr Ala Leu Ile Val Gly Ile Val Ala Gly Val Leu Gly Val Leu Leu  
 305 310 315 320  
 Ile Leu Gly Ala Val Val Ala Val Val Arg Arg Lys Lys His Ser Ser  
 325 330 335  
 Asp Gly Lys Gly Gly Arg Tyr Thr Pro Ala Ala Gly Gly His Arg Asp  
 340 345 350  
 Gln Gly Ser Asp Asp Ser Leu Met Pro  
 355 360

(2) INFORMATION FOR SEQ ID NO:23:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 365 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein.

## (ix) FEATURE:

- (A) NAME/KEY: Protein
- (B) LOCATION: 1..365
- (D) OTHER INFORMATION: /note= "Human Major Histocompatibility Class I (MHC) protein"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```

Met Ala Val Met Ala Pro Arg Thr Leu Val Leu Leu Leu Ser Gly Ala
1          5          10          15
Leu Ala Leu Thr Gln Thr Trp Ala Gly Ser His Ser Met Arg Tyr Phe
20          25          30
Phe Thr Ser Val Ser Arg Pro Gly Arg Gly Glu Pro Arg Phe Ile Ala
35          40          45
Val Gly Tyr Val Asp Asp Thr Gln Phe Val Arg Phe Asp Ser Asp Ala
50          55          60
Ala Ser Gln Arg Met Glu Pro Arg Ala Pro Trp Ile Glu Gln Glu Gly
65          70          75          80
Pro Glu Tyr Trp Asp Gly Glu Thr Arg Lys Val Lys Ala His Ser Gln
85          90          95
Thr His Arg Val Asp Leu Gly Thr Leu Arg Gly Tyr Tyr Asn Gln Ser
100         105         110
Glu Ala Gly Ser His Thr Leu Gln Met Met Phe Gly Cys Asp Val Gly
115         120         125
Ser Asp Trp Arg Phe Leu Arg Gly Tyr His Gln Tyr Ala Tyr Asp Gly
130         135         140
Lys Asp Tyr Ile Ala Leu Lys Glu Asp Leu Arg Ser Trp Thr Ala Ala
145         150         155         160
Asp Met Ala Ala Gln Thr Thr Lys His Lys Trp Glu Ala Ala His Val
165         170         175
Ala Glu Gln Leu Arg Ala Tyr Leu Glu Gly Thr Cys Val Glu Trp Leu
180         185         190
Arg Arg Tyr Leu Glu Asn Gly Lys Glu Thr Leu Gln Arg Thr Asp Ala
195         200         205
Pro Lys Thr His Met Thr His His Ala Val Ser Asp His Glu Ala Thr
210         215         220
Leu Arg Cys Trp Ala Leu Ser Phe Tyr Pro Ala Glu Ile Thr Leu Thr

```

|     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|
| 225 |     | 230 |     | 235 |     | 240 |
| Trp | Gln | Arg | Asp | Gly | Glu | Asp |
|     |     |     |     | 245 |     |     |
|     |     |     |     |     |     | 250 |
|     |     |     |     |     |     |     |
| Thr | Arg | Pro | Ala | Gly | Asp | Gly |
|     |     |     | 260 |     |     |     |
|     |     |     |     |     |     | 265 |
|     |     |     |     |     |     |     |
| Val | Pro | Ser | Gly | Gln | Glu | Gln |
|     |     | 275 |     |     |     |     |
|     |     |     |     |     |     | 280 |
|     |     |     |     |     |     |     |
| Gly | Leu | Pro | Lys | Pro | Leu | Thr |
|     | 290 |     |     |     |     | 295 |
|     |     |     |     |     |     |     |
|     |     |     |     |     |     | 300 |
|     |     |     |     |     |     |     |
| Thr | Ile | Pro | Ile | Val | Gly | Ile |
| 305 |     |     |     | 310 |     |     |
|     |     |     |     |     |     | 315 |
|     |     |     |     |     |     |     |
| Val | Ile | Thr | Gly | Ala | Val | Val |
|     |     |     | 325 |     |     |     |
|     |     |     |     |     |     | 330 |
|     |     |     |     |     |     |     |
| Ser | Asp | Arg | Lys | Gly | Gly | Ser |
|     |     |     | 340 |     |     |     |
|     |     |     |     |     |     | 345 |
|     |     |     |     |     |     |     |
|     |     |     |     |     |     | 350 |
|     |     |     |     |     |     |     |
| Ala | Gln | Gly | Ser | Asp | Val | Ser |
|     |     | 355 |     |     |     |     |
|     |     |     |     |     |     | 360 |
|     |     |     |     |     |     |     |
|     |     |     |     |     |     | 365 |

## (2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

ACATGGTTAA GGCCTGTTGC

20

## (2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GCCACATCTG GCTTGAAATT

20

## (2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (ix) FEATURE:
  - (A) NAME/KEY: modified\_base
  - (B) LOCATION: 1
  - (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-biotinylated adenine  
(bio-A)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

NGCTGTTTCGT GTTCTATGAT C

21

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (ix) FEATURE:
  - (A) NAME/KEY: modified\_base
  - (B) LOCATION: 1
  - (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-biotinylated adenine  
(bio-A)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

NGCTGTTTCGT GTTCTATGAT G

21

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

- (ix) FEATURE:
  - (A) NAME/KEY: modified\_base
  - (B) LOCATION: 1
  - (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-phosphorylated adenine"

(p-A)"

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 19
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 3'-digoxigenin-conjugated  
adenine (A-dig)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

NTGAGAGTCG CCGTGTGGN

19

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 39 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GGAAGAGCAG AGATATACGT GCCAGGTGGA GCACCCAGG

39

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 39 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GGAAGAGCAG AGATATACGT ACCAGGTGGA GCACCCAGG

39

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CAAAAGAAGC GGAGATTTAA CG

22

## (2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

AGATTTAACG GGGACGTGC

19

## (2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 22 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

AGAGGTCACA TGATGTGTCA CC

22

## (2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 19 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

AGGAGGCACT TGTGGTCC

19

## (2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 21 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:35:

AAAATCACAA CCACAGCAAA G

21

(2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:36:

TTCCACAGT GAGTCTGCAG

20

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CAATGGGGAT GGGACCTAC

19

(2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:38:

ATATACGTGC CAGGTGGAGC

20

(2) INFORMATION FOR SEQ ID NO:39:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CCTCTTCACA ACCCCTTTCA

20

(2) INFORMATION FOR SEQ ID NO:40:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

CATAGCTGTG CAACTCACAT CA

22

(2) INFORMATION FOR SEQ ID NO:41:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 40 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

AGCTGTTCGT GTTCTATGAT CATGAGAGTC GCCGTGTGGA

40

(2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 40 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

AGCTGTTCGT GTTCTATGAT GATGAGAGTC GCCGTGTGGA

40

(2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 32 base pairs  
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

TGTTCTATGA TCATGAGAGT CGCCGTGTGG AG

32

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 32 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

TGTTCTATGA TCATGAGTGT CGCCGTGTGG AG

32

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 17 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

ATATACGTGC CAGGTGG

17

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 17 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

ATATACGTAC CAGGTGG

17

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 17 base pairs  
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

TCTATGATCA TGAGAGT

17

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 17 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

TCTATGATCA TGAGAGT

17

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 17 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

TGGGTGCTCC ACCTGGC

17

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 17 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

TGGGTGCTCC ACCTGGT

17

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

CACACGGCGA CTCTCATG

18

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

CACACGGCGA CTCTCATC

18

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = fluorescein-labeled guanine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

NGAAGAGCAG AGATATACGT

20

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = fluorescein-labeled guanine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

NGCCTGGGTG CTCCACCTGG

20

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = fluorescein-labeled arginine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

NGCTGTTCGT GTTCTATGAT

20

(2) INFORMATION FOR SEQ ID NO:56:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = fluorescein-labeled cytosine"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

NTCCACACGG CGACTCTCAT

20

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

GCTCCACCTG GCACG

15

(2) INFORMATION FOR SEQ ID NO:58:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 15 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

GCTCCACCTG GTACG

15

(2) INFORMATION FOR SEQ ID NO:59:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 15 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

GCGACTCTCA TCATC

15

(2) INFORMATION FOR SEQ ID NO:60:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 15 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

GCGACTCTCA TGATC

15

(2) INFORMATION FOR SEQ ID NO:61:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: modified\_base

(B) LOCATION: 1  
(D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-biotinylated guanine  
(bio-G)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

NCCTGGGTGC TCCACCTGGC

20

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: modified\_base  
(B) LOCATION: 1  
(D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-phosphorylated adenine  
(p-A)"

(ix) FEATURE:

(A) NAME/KEY: modified\_base  
(B) LOCATION: 20  
(D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 3'-digoxigenin-conjugated  
cytosine (C-dig)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

NCGTATATCT CTGCTCTTCN

20

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: modified\_base  
(B) LOCATION: 1  
(D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-biotinylated guanine  
(bio-G)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

NAAGAGCAGA GATATACGTG

20

## (2) INFORMATION FOR SEQ ID NO:64:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-phosphorylated cytosine  
(p-C)"

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 20
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 3'-digoxigenin-conjugated  
cytosine (C-dig)"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

NCAGGTGGAG CACCCAGGCN

20

## (2) INFORMATION FOR SEQ ID NO:65:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-biotinylated guanine  
(bio-G)"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

NCCTGGGTGC TCCACCTGGT

20

## (2) INFORMATION FOR SEQ ID NO:66:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: modified\_base

(B) LOCATION: 1

(D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-phosphorylated adenine  
(p-A)"

(ix) FEATURE:

(A) NAME/KEY: modified\_base

(B) LOCATION: 20

(D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 3'-digoxigenin-conjugated  
cytosine (C-dig)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

NCGTATATCT CTGCTCTTCN

20

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: modified\_base

(B) LOCATION: 1

(D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-biotinylated guanine  
(bio-G)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

NAAGAGCAGA GATATACGTA

20

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: modified\_base

(B) LOCATION: 1

(D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-phosphorylated cytosine  
(p-C)"

(ix) FEATURE:

(A) NAME/KEY: modified\_base  
(B) LOCATION: 20  
(D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 3'-digoxigenin-conjugated  
cytosine (C-dig)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

NCAGGTGGAG CACCCAGGCN

20

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: modified\_base  
(B) LOCATION: 1  
(D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-biotinylated thymine  
(bio-T)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

NCCACACGGC GACTCTCATG

20

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: modified\_base  
(B) LOCATION: 1  
(D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-phosphorylated adenine  
(p-A)"

(ix) FEATURE:

(A) NAME/KEY: modified\_base  
(B) LOCATION: 20  
(D) OTHER INFORMATION: /mod\_base= OTHER

/note= "N = 3'-digoxigenin-conjugated  
thymine (T-dig)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

NTCATAGAAC ACGAACAGCN

20

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-biotinylated guanine  
(bio-G)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

NCTGTTCTG TTCTATGATC

20

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-phosphorylated adenine  
(p-A)"

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 20
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 3'-digoxigenin-conjugated  
guanine (G-dig)"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

NTGAGAGTCG CCGTGTGGAN

20

## (2) INFORMATION FOR SEQ ID NO:73:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-biotinylated thymine  
(bio-T)"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

NCCACACGGC GACTCTCATC

20

## (2) INFORMATION FOR SEQ ID NO:74:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-phosphorylated adenine  
(p-A)"

## (ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 20
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 3'-digoxigenin-conjugated  
thymine (T-dig)"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

NTCATAGAAC ACGAACAGCN

20

## (2) INFORMATION FOR SEQ ID NO:75:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-biotinyllaced guanine  
(bio-G) "

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

NCTGTTCTG TG TTCTATGATG

20

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 1
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 5'-phosphorylated adenine  
(p-A) "

(ix) FEATURE:

- (A) NAME/KEY: modified\_base
- (B) LOCATION: 20
- (D) OTHER INFORMATION: /mod\_base= OTHER  
/note= "N = 3'-digoxigenin-conjugated  
guanine (G-dig) "

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

NTGAGAGTCG CCGTGTGGAN

20